Particle Filters for Dual State and Parameter Estimation of Nonlinear Systems with Application to Fault Diagnosis of Gas Turbine Engines *

Najmeh Daroogheh[1], Nader Meskin[2] and Khashayar Khorasani[1]

Abstract

In this paper, a novel dual estimation methodology is developed for both time-varying parameters and states of a nonlinear stochastic system based on the Recursive Prediction Error (RPE) concept and Particle Filtering (PF) scheme. Our developed methodology in based on a parallel implementation of state and parameter estimation filters as opposed to using a single filter for estimating the augmented states and parameters. The convergence and stability of our proposed dual estimation strategy are shown formally to be guaranteed under certain conditions. The proposed dual estimation framework is then utilized for addressing the challenging problem of fault diagnosis of nonlinear systems. The performance capabilities of our proposed fault diagnosis methodology is demonstrated and evaluated by its application to a gas turbine engine through accomplishing state and parameter estimation under simultaneous component fault scenarios. The health parameters of the system are considered to be slowly time-varying in the steady state condition of the engine operation. Extensive simulation results are provided to substantiate and justify the superiority of our proposed fault diagnosis methodology when compared to two other alternative diagnostic techniques that are available in the literature.

I. INTRODUCTION

State estimation of systems is a fundamental problem in control, signal processing, and fault diagnosis fields [1]. Investigations on both linear and nonlinear state estimation and filtering in

*This publication was made possible by NPRP grant No. 4-195-2-065 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

[1]N. Daroogheh and K. Khorasani are with the Department of Electrical and Computer Engineering, Concordia University, Montreal, Quebec, Canada n_daroog@encs.concordia.ca and kash@ece.concordia.ca

[2]N. Meskin is with the Department of Electrical Engineering, Qatar University, Doha, Qatar nader.meskin@qu.edu.qa

stochastic environments have been an active area of research during the past several decades. Linear state estimation methods use a simpler representation of an actual nonlinear system and can provide an acceptable performance locally around an operating point and in the steady state operational condition of the system. However, as nonlinearities in the system dynamics become dominant, performance of linear approaches deteriorates and linear algorithms will not necessarily converge to an accurate solution. Although an optimal state estimation solution for linear filtering methods exists, nonlinear filtering methods suffer from sub-optimal or near-optimal solutions. Consequently, investigation of nonlinear estimation and filtering problems remain a challenging research in the above fields.

Numerous studies have been conducted in the literature to solve and analyze standard nonlinear filtering problems [2]–[8]. These methods can be broadly categorized into [7]: (a) linearization methods (extended Kalman filters (EKF)) [2], (b) approximation methods using finite-dimensional nonlinear filters [3], (c) particle filter (PF) methods as one of the most popular Bayesian recursive methods for state estimation [4], (d) classical partial differential equation (PDE) methods for approximating a solution to the Zakai equation [5], (e) Wiener chaos expansion methods [6], (f) moment methods [7], and (g) high dimensional nonlinear Kalman filter methods known as the Cubature Kalman filters [8].

Estimation of states and parameters of nonlinear systems have been tackled by using extended Kalman filtering (EKF) methods in [9], where the accuracy of parameter estimates is only validated through experimental results. An off-line parameter estimation method was developed in [10], in which the effects of each parameter on the system measurements were quantified by using the principal component analysis (PCA). However, this method is inherently off-line and cannot be used to track on-line variations of the system parameters. Dual state/parameter estimation methods have also been developed for nonlinear systems based on (i) EKF in [11], (ii) unscented Kalman filters (UKF) in [12] for slowly-varying parameters, and (iii) ensemble Kalman filters in [13] for static (fixed or constant) parameters.

While dual estimation methods based on UKF achieve more accurate results as compared to the EKF-based dual methods, both of these methods are applicable to systems that are considered to be affected by the process and measurement noise with Gaussian distributions. Therefore, particle filter-based dual state/parameter estimation methods are introduced as powerful tools for nonlinear systems that are affected by non-Gaussian process and measurement noise [4], [14].

In cases where the fixed or constant system parameters are unknown and need to be estimated from either on-line or off-line approaches, two main classes of estimation methods have been proposed in the literature based on the numerical particle filtering (PF) and the maximum likelihood filtering [15]. These methods are mainly developed to only estimate fixed (constant) parameters of a system. A combination of particle filtering and gradient algorithm based on the stochastic approximation technique was utilized in [16] to develop an adaptive fixed parameter estimation method. The efficiency of this method was assessed to be more reliable than general particle filtering methods that are developed for parameter estimation problems.

In the framework of particle filters, a general method that is capable of *simultaneously* estimating the static (constant or fixed) parameters and the time-varying states of a system is developed in [17]. This work is based on the sequential Monte Carlo (SMC) method in which an artificial dynamic evolution model is considered for the unknown model parameters. In order to overcome the degeneracy concerns due to the particle filtering, kernel smoothing technique [18] as a method for smoothing the approximation of the parameters conditional density was invoked. The estimation algorithm was further improved by re-interpretation of the artificial evolution algorithm according to the shrinkage scaling concept. However, the proposed method in [18] is only applicable for estimating fixed parameters of the system and they used the augmented state/parameter vector for the purpose of system estimation.

In the present paper, we propose to extend the particle filtering approach for state estimation by utilizing the recursive prediction error (RPE) scheme to estimate the slowly time-varying parameters in a dual state/parameter estimation filter as opposed to an augmented approach that was introduced in [17]. The RPE method is used for both off-line and on-line system identification by utilizing a quadratically convergent scheme [19], [20]. The concept of the kernel mean shrinkage [18] is also used in our proposed scheme to avoid overdispersion in the variance of the estimated parameters.

Towards the above end, the main contribution of this paper is to utilize nonlinear Bayesian and Sequential Monte Carlo (SMC) methods to develop a unified framework for both the state and parameter estimation problems. Our methodology is based on solving the Bayesian recursive relations by using SMC methods. An on-line parameter estimation scheme is developed based on the recursive prediction error (RPE) method [19] by using the particle filter (PF) approach. Specifically, by using the prediction error to correct for the changes in the system parameters,

a novel method is proposed for parameter estimation of nonlinear systems based on the PF. In the implementation of the algorithm, a dual structure is proposed for both state and parameter estimation in the PF framework. The hidden states, and the variations in the system parameters are estimated through two parallel filters.

The other equally important problem that is considered in this paper deals with fault detection and isolation (FDI) of nonlinear systems, also known as the fault diagnosis (FD) problem. Proper utilization of FDI systems can improve the reliability and operational safety of industrial systems. The FDI techniques of interest in this work use the concept of analytical redundancy [21], thereby not requiring any additional and redundant hardware. However, these analytical methods are more challenging to develop due to requiring properties such as robustness against noise, model uncertainties and unknown disturbances [22], [23].

State estimation methods have been applied for development of model-based fault diagnosis (FD) and health monitoring techniques in recent years [24]. In these methods, accuracy of the considered dynamical model plays an important and crucial role in the performance of the developed FD scheme. Therefore, methods that are designed by explicitly taking into account the nonlinear dynamical characteristics of the system are considered to yield more accurate solutions to the fault diagnosis problem [25].

Fault detection and isolation is achieved by using the concatenated wavelet transform variances and discriminate analysis in [26]. This method was proposed as an alternative to the multiple-model approach used for FDI. It was shown that it can improve the overall performance in fault classification of mechanical systems, however it is applicable to systems that are affected by the stationary noise processes. A fault detection and diagnosis scheme (FDD) for continuous-time stochastic dynamic systems with time delays and non-Gaussian variables is proposed in [27] based on an observer-based optimal FDD using the conditional probability distributions. However, in this method it is assumed that the output pdfs can be approximated by using the square root B-spline models.

Component fault diagnosis of dynamical systems can be achieved through dual state/parameter estimation methods [24], [28] by first detecting the faulty competent and then by determining its severity and isolating its location. In this work our proposed dual state and parameter estimation strategy is applied for component fault diagnosis of a gas turbine engine based on the nonlinear gas path analysis approach [29]. The component fault diagnosis in gas turbine

engines is addressed by utilizing a multiple-model approach in [29], [30]. Whereas the multiple-model approach can detect and isolate faults based on the predefined faulty models, utilization of the parameter estimation scheme in our developed dual state/parameter estimation enables the diagnostic algorithm to detect, isolate and accurately identify the type and severity of simultaneous fault scenarios in the gas turbine system.

The remainder of this paper is organized as follows. In Section II, the statement of the nonlinear filtering problem is introduced. Our main proposed dual state/parameter estimation scheme is developed in Section III, in which state and parameter estimation methods are first developed in parallel and separately and subsequently integrated together for simultaneously estimating the system state/parameters. The stability and convergence properties of the proposed scheme under certain conditions are also provided in Section III. In Section IV, extensive simulation results and case studies are provided to demonstrate and justify the merits of our proposed method for fault diagnosis of a gas turbine engine under simultaneous component faults. Finally, the paper is concluded in Section V.

II. PROBLEM STATEMENT

The problem under consideration is to obtain an optimal estimate of the states as well as the time-varying parameters of a nonlinear system whose dynamics is governed by a discrete-time stochastic model given by,

$$x_{t+1} = f_t(x_t, \theta_t, \omega_t), \tag{1}$$

$$y_t = h_t(x_t, \theta_t) + \nu_t, \tag{2}$$

where $f_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_\omega} \longrightarrow \mathbb{R}^{n_x}$ is a known nonlinear function that specifies the state at the next time step $t \in \mathbb{N}$, $\theta_t \in \mathbb{R}^{n_\theta}$ is the unknown and possibly time-varying parameter vector at time t that is governed by an unknown dynamics. The function $h_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \longrightarrow \mathbb{R}^{n_y}$ is a known nonlinear function representing the map between the states, parameters and the system measurements at time t, ω_t and ν_t are uncorrelated stochastic process and measurement noise sequences with covariance matrices Q_t and R_t, respectively.

Our main goal in the dual state and parameter estimation problem is to approximate the following conditional expectations:

$$\mathbb{E}(\phi_1(x_t)|y_{1:t}, \theta_{t-1}) = \int \phi_1(x_t) p(x_t|y_{1:t}, \theta_{t-1}) \mathrm{d}x_t, \tag{3a}$$

$$\mathbb{E}(\phi_2(\theta_t)|y_{1:t}, x_t) = \int \phi_2(\theta_t) p(\theta_t|y_{1:t}, x_t) \mathrm{d}\theta_t, \tag{3b}$$

where $y_{1:t} = (y_1, y_2, ..., y_t)$ denotes the available observations up to time t, and $\phi_1 : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$, and $\phi_2 : \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}$ are functions of states and parameters, respectively, that are to be estimated. The conditional probability functions $p(x_t|y_{1:t}, \theta_{t-1})\mathrm{d}x_t$ and $p(\theta_t|y_{1:t}, x_t)\mathrm{d}\theta_t$ are to be approximated by the designed particle filters (PFs) through determining the filtering distributions according to

$$\begin{aligned} \hat{p}_N(x_t|y_{1:t}, \theta_{t-1})\mathrm{d}x_t &= \sum_{i=1}^{N} w_{x_t}^{(i)} \delta_{x_t^{(i)}}(\mathrm{d}x_t), \\ \hat{p}_M(\theta_t|y_{1:t}, x_t)\mathrm{d}\theta_t &= \sum_{j=1}^{M} w_{\theta_t}^{(j)} \delta_{\theta_t^{(j)}}(\mathrm{d}\theta_t), \end{aligned} \tag{4}$$

where the subscripts N and M imply that the state and parameter conditional probability distributions are obtained from N and M particles, respectively. Each particle $x_t^{(i)}$ has a weight $w_{x_t}^{(i)}$ and each particle $\theta_t^{(j)}$ has a related weight $w_{\theta_t}^{(j)}$, where $\delta(.)$ denotes the Dirac-delta function mass that is positioned at x_t or θ_t.

Based on the approximation that is used in the representations in equation (4), our goal is to address the convergence properties of the to be designed estimators to their true optimal estimates and also to develop and demonstrate under what conditions this convergence remains valid. Subsequently in Section IV, the models (1) and (2) are used to address the problem of fault diagnosis of a dynamical nonlinear system when the system parameters are considered to be affected by a multiplicative fault vector parameter.

III. Proposed Dual State/Parameter Estimation Framework

In this section, the main theoretical framework of our proposed dual state/parameter filtering of the nonlinear system (1) and (2) is introduced and developed.

A. Dynamic Model in Presence of Time-Varying Parameters

Our first task is to represent the model (1) and (2) into a different framework for our subsequent theoretical developments. Let $(\Omega, \mathcal{F}, P)$ denote the probability space on which the three real vector-valued stochastic processes, $X = \{X_t, t = 1, 2, ...\}, \Theta = \{\Theta_t, t = 1, 2, ...\}$ and $Y = \{Y_t, t = 1, 2, ...\}$ are defined. The n_x-dimensional process X describes the evolution of the hidden

states, the n_θ-dimensional process Θ describes the evolution of the system hidden parameters which are conditionally independent of the states, and the n_y-dimensional process Y denotes the observation process of the system.

The processes X and Θ are Markov processes with the associated initial state and initial parameter X_0 and Θ_0, respectively. They are drawn from the initial distributions $\pi_{x_0}(\mathrm{d}x_0)$ and $\pi_{\theta_0}(\mathrm{d}\theta_0)$, respectively. The dynamic evolution of the states and parameters are modeled by the Markov transition kernels $K_x(\mathrm{d}x_t|x_{t-1},\theta_{t-1})$ and $K_\theta(\mathrm{d}\theta_t|\theta_{t-1},x_t)$ that also admit densities with respect to the Lebesgue measure [1], such that

$$P(X_t \in A_1|X_{t-1}=x_{t-1},\Theta_{t-1}=\theta_{t-1}) = \int_{A_1} K_x(x_t|x_{t-1},\theta_{t-1})\mathrm{d}x_t, \tag{5}$$

$$P(\Theta_t \in A_2|\Theta_{t-1}=\theta_{t-1},X_t=x_t) = \int_{A_2} K_\theta(\theta_t|\theta_{t-1},x_t)\mathrm{d}\theta_t, \tag{6}$$

for all $A_1 \in \mathcal{B}(\mathbb{R}^{n_x})$ and $A_2 \in \mathcal{B}(\mathbb{R}^{n_\theta})$, where $\mathcal{B}(\mathbb{R}^{n_x})$ and $\mathcal{B}(\mathbb{R}^{n_\theta})$ denote the Borel σ-algebra on $\mathbb{R}^{n_x}$ and $\mathbb{R}^{n_\theta}$, respectively. The transition kernel $K_x(x_t|x_{t-1},\theta_{t-1})$ is a probability distribution function (pdf) that follows the pdf of the stochastic process in equation (1). The probability density function for approximating the parameter kernel transition $K(\theta_t|\theta_{t-1},x_t)$ is to be provided in the next subsequent subsections.

Given the states and parameters, the observations Y_t are conditionally independent and have the marginal distribution with a density with respect to the Lebesgue measure as given by,

$$P(Y_t \in B|X_t=x_t,\Theta_t=\theta_t) \quad = \int_B \rho(y_t|x_t,\theta_t)\mathrm{d}y_t, \tag{7}$$

where $\rho(y_t|x_t,\theta_t)$ is a probability density function which follows the probability density function of the stochastic process in equation (2).

In the dual state/parameter estimation framework at first the state x_t is estimated (which is denoted by $\hat{x}_{t|t}$) and this estimated value at time t is then used to estimate the parameter value θ_t at time t (which is denoted by $\hat{\theta}_{t|t}$). In many practical situations, the evolution of the parameters are not specified *a priori*, therefore it is necessary to consider a given evolution for the parameters in order to design the estimation filter. In our proposed dual structure for the state estimation filter, first the parameters are assumed to be constant at $t-1$ at their estimated value $\hat{\theta}_{t-1|t-1}$, and then for the parameter estimation filter they are evolved to the next time instant by applying

[1]The transition kernel $K(\mathrm{d}x_t|x_{t-1})$ admits density with respect to the Lebesgue measure if one can write $P(X_t \in \mathrm{d}x_t|X_{t-1} = x_{t-1}) = K(\mathrm{d}x_t|x_{t-1}) = K(x_t|x_{t-1})\mathrm{d}x_t$.

an update law based on the recursive prediction error (RPE) method. The details regarding our proposed methodology are presented in the subsequent subsections in which the filtering of the states and parameters are described and developed.

B. The State Estimation Problem

For designing both the state and parameter estimation filters, our main goals are to approximate the integrals in equations (5) and (6) by invoking the particle filter scheme and to approximate the estimate of the conditional state and parameter distributions. Consider $\pi_{x_{t|t-1}}(\mathrm{d}x_t) = \int_{\mathbb{R}^{n_x}} \pi_{x_{t-1|t-1}}(\mathrm{d}x_{t-1})K_x(\mathrm{d}x_t|x_{t-1}, \theta_{t-1})$ as the *a priori* state estimation distribution before the observation at time t becomes available, and $\pi_{\theta_{t-1|t-1}}(\mathrm{d}\theta_{t-1})$ as the conditional distribution of the parameter at time $t-1$. The *a posteriori* state distribution after the observation at the time instant t becomes available is obtained according to the following rule,

$$\pi_{x_{t|t}}(\mathrm{d}x_t) \propto \rho(y_t|x_t, \theta_{t-1})\pi_{x_{t|t-1}}(\mathrm{d}x_t)\pi_{\theta_{t-1|t-1}}(\mathrm{d}\theta_{t-1}). \tag{8}$$

For our to be designed state estimation filter the parameters are frozen and fixed to their estimated values at the previous time step, i.e. $\hat{\theta}_{t-1|t-1}$, hence it is assumed that $\hat{\theta}_{t-1|t-1}$ is known for this filter. Therefore, the last term in the right hand side of equation (8) is set to one.

The particle filter (PF) procedure for implementation of the state estimation and for determining $\pi_{x_{t|t}}(\mathrm{d}x_t)$ consists of two main steps, namely (a) the prediction step, and (b) the measurement update step (time update step). Consider one is given N particles at time t. The prediction step utilizes the knowledge of the previous distribution of the states as well as the previous parameter estimate, that are denoted by $\{\hat{x}^{(i)}_{t-1|t-1},\ i = 1, ..., N\}$ (corresponding to N estimated state particles that follow the distribution $\pi_{x_{t-1|t-1}}(\mathrm{d}x_{t-1})$) and $\hat{\theta}_{t-1|t-1}$, respectively, and the process model as given by equation (1). In other words, the prediction step is explicitly governed by the following equations for $i = 1, ..., N$, as

$$\hat{x}^{(i)}_{t|t-1} = f_t(\hat{x}^{(i)}_{t-1|t-1}, \hat{\theta}_{t-1|t-1}, \omega^{(i)}_t) \tag{9a}$$

$$\hat{y}^{(i)}_{t|t-1} = h_t(\hat{x}^{(i)}_{t|t-1}, \hat{\theta}_{t-1|t-1}), \tag{9b}$$

$$\Sigma_{\hat{x}_{t|t-1}} = (\hat{x}^{(i)}_{t|t-1} - \frac{1}{N}\sum_{i=1}^{N}\hat{x}^{(i)}_{t|t-1})(\hat{x}^{(i)}_{t|t-1} - \frac{1}{N}\sum_{i=1}^{N}\hat{x}^{(i)}_{t|t-1})^{\mathrm{T}}, \tag{9c}$$

where $\omega^{(i)}_t$ denotes the process noise related to each particle $\hat{x}^{(i)}_{t|t-1}$ and is drawn from the noise distribution with the probability distribution function $p_{\omega_t}(.)$, and $\hat{x}^{(i)}_{t|t-1}$ denotes the independent

samples generated from equation (9a) (with the transition kernel as given by $K_x(.)$) through the N particles. The initial covariance of the noise term is chosen less than or equal to the one in equation (1). Moreover, $\hat{y}^{(i)}_{t|t-1}$ denotes the independent samples of the predicted outputs evaluated at $\hat{x}^{(i)}_{t|t-1}$ samples, and $\Sigma_{\hat{x}_{t|t-1}}$ denotes the *a priori* state estimation covariance matrix.

For the first step, the one-step ahead prediction distribution known as the *a priori* state estimation distribution is now given by,

$$\tilde{\pi}^N_{x_{t|t-1}}(\mathrm{d}x_t) \triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{\hat{x}^{(i)}_{t|t-1}}(\mathrm{d}x_t), \tag{10}$$

For the second step, the information on the present observation y_t is used. This results in approximating $\pi_{x_{t|t}}(\mathrm{d}x_t)$, where $\hat{\theta}_{t-1|t-1}$ is considered as given from a parameter estimation filter and obtained from the distribution $\pi^M_{\theta_{t-1|t-1}}(\mathrm{d}\theta_{t-1})$, where $M \geq N$ is the number of particles in the parameter estimation filter (details are given in the Subsection III-C), and N is the minimum number of the required particles for the state estimator to converge. Specifically, according to the Bayes' rule, we have the approximation:
$\tilde{\pi}^N_{x_{t|t}}(\mathrm{d}x_t) \propto \frac{\rho(y_t|x_t,\theta_{t-1})\tilde{\pi}^N_{x_{t|t-1}}(\mathrm{d}x_t)\pi^M_{\theta_{t-1|t-1}}(\mathrm{d}\theta_{t-1})}{\int_{\mathbb{R}^{n_x}}\int_{\mathbb{R}^{n_\theta}}\rho(y_t|x_t,\theta_{t-1})\tilde{\pi}^N_{x_{t|t-1}}(\mathrm{d}x_t)\pi^M_{\theta_{t-1|t-1}}(\mathrm{d}\theta_{t-1})}$, since the parameter estimate at $t-1$, i.e. $\hat{\theta}_{t-1|t-1}$ is assumed to be known and approximated from the mean of the *a posteriori* parameter estimate distribution at $t-1$. Consequently, $\tilde{\pi}^N_{x_{t|t}} \propto \frac{\rho(y_t|\hat{x}^{(i)}_{t|t-1},\hat{\theta}_{t-1|t-1})\delta_{\hat{x}^{(i)}_{t|t-1}}(\mathrm{d}x_t)}{\frac{1}{N}\sum_{i=1}^{N}\rho(y_t|\hat{x}^{(i)}_{t|t-1},\hat{\theta}_{t-1|t-1})}$.

Remark 1. *In order for the optimal state filter to exist, as per existence of the probability distribution function in [31], one requires that* $\frac{1}{N}\sum_{i=1}^{N}\rho(y_t|\hat{x}^{(i)}_{t|t-1},\hat{\theta}_{t-1|t-1}) > 0$. *Hence, in the state estimation filter of the dual structure one must ensure that* $\frac{1}{N}\sum_{i=1}^{N}\rho(y_t|\hat{x}^{(i)}_{t|t-1},\hat{\theta}_{t-1|t-1}) \geq r^1_{th}$, r^1_{th} *denotes a threshold that is selected for sufficiently large* N *and* $\hat{\theta}_{t-1|t-1}$ *is the mean of the a posteriori distribution of* $\hat{\theta}^{(j)}_{t-1|t-1}$ *denoted by* $\pi^M_{\theta_{t-1|t-1}}(\mathrm{d}\theta_{t-1})$ *and approximated by the* $j = 1, ..., M$ *particles.*

It follows that for the *a priori* state estimation algorithm (9a)-(9c) to proceed, the condition in Remark 1 must be satisfied, otherwise a new particle population has to be generated and this condition is checked again [31]. A suitable initial choice for r^1_{th} can be taken as the first standard deviation of the acceptable normalized error in the output measurement (the signal y_t in equation (2)) in the steady state mode of the system operation. The condition stated in Remark 1 is invoked to reduce the risk of the filter divergence according to [31].

The particle weights $w^{(i)}_{x_t}$ are updated by the likelihood function (the importance function)

according to $w_{x_t}^{(i)} \sim p_{\nu_t}(y_t - \hat{y}_{t|t-1}^{(i)}) = \rho(y_t|\hat{x}_{t|t-1}^{(i)}, \hat{\theta}_{t-1|t-1})$, where $p_{\nu_t}(.)$ denotes the probability distribution function of the additive noise of the output and is evaluated at $y_t - \hat{y}_{t|t-1}^{(i)}$. Next, a resampling is performed to generate N new particles as $\{\bar{x}_{t|t-1}^{(i)}\}_{i=1}^{N}$ with replacements from $\{\hat{x}_{t|t-1}^{(i)}\}_{i=1}^{N}$, where the probability for taking the sample k is $P(\bar{x}_{t|t-1}^{(i)} = \hat{x}_{t|t-1}^{(k)}) = \tilde{w}_{x_t}^{(k)} \triangleq \frac{\rho(y_t|\hat{x}_{t|t-1}^{(k)}, \hat{\theta}_{t-1|t-1})}{\sum_{k=1}^{N} \rho(y_t|\hat{x}_{t|t-1}^{(k)}, \hat{\theta}_{t-1|t-1})}$, and where $\tilde{w}_{x_t}^{(k)}$ for $k = 1, ..., N$ denotes the normalized particle weights. In other words, the particle selection in the resampling step is performed for the particles with higher probabilities of $\rho_{\nu_t}(y_t - \hat{y}_{t|t-1}^{(k)})$.

Although the total number of the particles after the resampling is the same as that before the resampling, certain particles for which the weight $\tilde{w}_{x_t}^{(i)}$ is greater than a threshold will be applied to their neighbors with lower weight values. After the resampling all the new particles are assigned equal weights as $\frac{1}{N}$. There are different methods for performing the resampling [4]. In this work, we have chosen the regularized particle filters (RPF) since they are capable of transforming the discrete-time approximation of the *a posteriori* state estimation distribution $\pi_{x_{t|t}}^{N}(\mathrm{d}x_t)$ into a continuous-time one. Consequently, the resampling step is modified in such a manner so that the new resampled particles are obtained from an absolutely continuous-time distribution with N different locations $\bar{x}_{t|t-1}^{(i)}$ from that of $\hat{x}_{t|t-1}^{(i)}$ [32]. Therefore, the *a posteriori* state estimation distribution is approximated by $\tilde{\pi}_{x_{t|t}}^{N}(\mathrm{d}x_t)$ before performing the resampling by using the reguralized particle filter (RPF) structure [32], and by $\pi_{x_{t|t}}^{N}(\mathrm{d}x_t)$ after performing the resampling as provided below,

$$
\begin{aligned}
&\tilde{\pi}_{x_{t|t}}^{N}(\mathrm{d}x_t) \approx \sum_{l=1}^{N_{reg}} \sum_{i=1}^{N} \tilde{w}_{x_t}^{(i)} \frac{|A^{-1}|}{b^{n_x}} \mathcal{K}(\frac{1}{b} A^{-1}(x_t^{reg_l} - (\hat{x}_{t|t-1}^{(i)}))),\\
&\tilde{w}_{x_t}^{(i)} \triangleq \frac{\rho(y_t|\hat{x}_{t|t-1}^{(i)}, \hat{\theta}_{t-1|t-1})}{\sum_{i=1}^{N} \rho(y_t|\hat{x}_{t|t-1}^{(i)}, \hat{\theta}_{t-1|t-1})},\\
&\pi_{x_{t|t}}^{N}(\mathrm{d}x_t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\bar{x}_{t|t-1}^{(i)}}(\mathrm{d}x_t) \to \hat{x}_{t|t} = \frac{1}{N} \sum_{i=1}^{N} \bar{x}_{t|t-1}^{(i)},
\end{aligned}
\tag{11}
$$

where $x_t^{reg_l},\ l = 1, .., N_{reg}$ denotes the regularized state vector evaluated at N_{reg} points which are obtained from the absolutely continuous-time distribution of the particles between the first standard deviation of the maximum value of the sampled particles $\hat{x}_{t|t-1}^{(i)}$ and their minimum value that is divided into N_{reg} desired regularized points. Hence, $\{\bar{x}_{t|t-1}^{(i)}\}_{i=1}^{N}$ is obtained from the continuous-time distribution through the regularization kernel $\mathcal{K}$ which is considered to be a symmetric density function on $\mathbb{R}^{n_x}$ [32]. The matrix A is chosen to yield a unit covariance value

in the new $\bar{x}^{(i)}_{t|t-1}$ population and $AA^{\mathrm{T}} = \Sigma_{\hat{x}_{t|t-1}}$. The constant b denotes the optimal bandwidth of the kernel, and $\hat{x}_{t|t}$ denotes the *a posteriori* state estimation at time t.

We are now in a position to introduce our overall particle filter (PF) scheme for implementing the state estimation filter. Our goal for proposing this algorithm is to ensure that an approximation to $\mathbb{E}(\phi(x_t)|y_{1:t}, \theta_{t-1})$ by $\phi(x_t) = x_t$ takes $\hat{x}_{t|t} \sim \pi^N_{x_{t|t}}(\mathrm{d}x_t) = \frac{1}{N}\sum_{i=1}^N \delta_{\hat{x}^{(i)}_{t|t}}(\mathrm{d}x_t)$, where $\pi^N_{x_{t|t}}(\mathrm{d}x_t)$ denotes the *a posteriori* distribution of $\{\bar{x}^{(i)}_{t|t-1}\}_{i=1}^N$ (after the resampling from $\{\hat{x}^{(i)}_{t|t-1}\}_{i=1}^N$), as given by $\hat{x}_{t|t} = \frac{1}{N}\sum_{i=1}^N \bar{x}^{(i)}_{t|t-1}$. The estimated output from the state estimation filter is also given by $\hat{y}_t = h_t(\hat{x}_{t|t}, \hat{\theta}_{t-1|t-1})$.

<u>*The State Estimation Particle Filter Scheme*</u>

1) Initialize the N particles, $\{x_0^{(i)}\}_{i=1}^N \sim \pi_{x_0}(\mathrm{d}x_0)$ and parameters with θ_0 (the mean of the parameter initial distribution $\pi_{\theta_0}(\mathrm{d}\theta_0)$).
2) Draw $\omega_t^{(i)} \sim p_{\omega_t}(.)$, where $p_{\omega_t}(.)$ denotes a given distribution for the process noise in the filter, and then predict the state particles $\hat{x}^{(i)}_{t|t-1}$ according to equation (9a).
3) If the condition in Remark 1 holds, that is $\frac{1}{N}\sum_{i=1}^N \rho(y_t|\hat{x}^{(i)}_{t|t-1}, \hat{\theta}_{t-1|t-1}) \geq r^1_{th}$ go to Step 4, otherwise return to Step 2.
4) Compute $\hat{y}^{(i)}_{t|t-1}$ from equation (9b) to obtain the importance weights $\{w^{(i)}_{x_t}\}_{i=1}^N$ as $w^{(i)}_{x_t} = \rho(y_t|\hat{x}^{(i)}_{t|t-1}, \hat{\theta}_{t-1|t-1}),\ i = 1, ..., N$, and normalize them to $\tilde{w}^{(i)}_{x_t} = \frac{w^{(i)}_{x_t}}{\sum_{i=1}^N w^{(i)}_{x_t}}$.
5) Resampling: Draw N new particles with the replacement for each $i = 1, ..., N$, according to $P(\bar{x}^{(i)}_{t|t-1} = \hat{x}^{(k)}_{t|t-1}) = \tilde{w}^{(k)}_{x_t},\ k = 1, ..., N$, from the regularized kernel $\mathcal{K}$ where $\bar{x}^{(i)}_{t|t-1} \sim \tilde{\pi}^N_{x_{t|t}}(\mathrm{d}x_t)$ as given by equation (11).
6) Calculate $\hat{x}_{t|t}$ from the conditional distribution given by equation (11), $\pi^N_{x_{t|t}}(\mathrm{d}x_t) = \frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}^{(i)}_{t|t-1}}(\mathrm{d}x_t)$ with equally weighted $\bar{x}^{(i)}_{t|t-1}$ as $\hat{x}_{t|t} = \frac{1}{N}\sum_{i=1}^N \bar{x}^{(i)}_{t|t-1}$.
7) Update the parameters from the parameter estimation filter (to be specified in the next subsection).
8) Set $t := t + 1$ and go to Step 2.

Following the implementation of the above state estimation filter, the parameter estimation filter that is utilized for adjusting the parameters is now described in detail.

C. The Parameter Estimation Problem

One of the main contributions of this paper is to develop a novel PF-based parameter estimation filter within our proposed dual state/parameter estimation framework by utilizing the recursive

prediction error (RPE) concept. For this part of the methodology, it is assumed that the parameters are not known *a priori* and are also time-varying. Moreover, the estimated states obtained from the state estimation filter provided in the previous subsection will be used. In other words, in the parameter estimation filter the estimated states are used as $x_t \approx \hat{x}_t = \hat{x}_{t|t}$. Therefore, it is imperative that one considers a dynamical model corresponding to the parameters evolution in order to estimate the density function $\pi_{\theta_{t|t}}(\mathrm{d}\theta_t)$.

The most common dynamical model that is considered for parameter propagation (in case of fixed system parameters) is the conventional artificial evolution law in which small random disturbances are added to the state particles (parameters) between the time steps [17]. However, in the present work, the conventional artificial evolution update law for the parameters is modified to include the output prediction error as a term allowing one to deal with the time variations in the parameters that can affect the system output.

In order to derive an update law for the parameters, an algorithm based on the recursive prediction error (RPE) method is proposed by minimizing the expectation of a quadratic function $\bar{J}(\theta_{t-1})$ with respect to θ_{t-1}. The considered cost function is to be minimized with respect to θ_{t-1}, since our parameter estimation algorithm for finding the distribution of the *a posteriori* parameter estimate is based on the kernel smoothing using the shrinkage of the particle locations. This method attempts to force the particles towards their mean value in the previous time step, i.e. the estimated value of θ_{t-1} that is denoted by $\hat{\theta}_{t-1|t-1}$ (before adding noise to the particles), which is also used in the state estimation filter for approximating $\hat{x}_{t|t}$. Therefore, our goal is to investigate the convergence properties of $\hat{\theta}_{t-1|t-1}$ whose boundedness ensures the boundedness of $\hat{\theta}_{t|t}$. Towards this end, the cost function is now selected as $\mathbb{E}(\bar{J}(\theta_{t-1})|y_{1:t-1}, x_t) = \int \bar{J}(\theta_{t-1})p(\theta_{t-1}|y_{1:t-1}, x_t)d\theta_{t-1}$, where the integral is approximated in the PF by $\mathbb{E}(\bar{J}(\theta_{t-1})|y_{1:t-1}, x_t) \approx \frac{1}{M}\sum_{j=1}^{M} \bar{J}(\hat{\theta}^{(j)}_{t-1|t-1})$.

The function $\bar{J}(\hat{\theta}^{(j)}_{t-1|t-1})$ now represents a quadratic function of the output prediction error related to each particle $j,\ j = 1, ..., M$. The prediction error is now defined by $\epsilon(t, \hat{\theta}^{(j)}_{t-1|t-1}) \triangleq \epsilon^{(j)}_t = y_t - h_t(\hat{x}_{t|t}, \hat{\theta}^{(j)}_{t-1|t-1})$, and $\hat{\theta}^{(j)}_{t-1|t-1}$ denotes the particle related to the estimated value of the parameter whose true value is denoted by $\theta^{\star}_{t-1}$ and which is clearly assumed to be unknown. Therefore, we define $\bar{J}(\hat{\theta}^{(j)}_{t-1|t-1}) = \frac{1}{\kappa}\sum_{\tau=t-\kappa}^{\tau=t} \mathbb{E}(l(\epsilon(\tau, \hat{\theta}^{(j)}_{t-1|t-1})))$ in which the expectation is taken over the observation sequence of κ samples. Let us now select the quadratic criterion $l(\epsilon(t, \hat{\theta}^{(j)}_{t-1|t-1}))$ as

$$l(\epsilon(t,\hat{\theta}^{(j)}_{t-1|t-1})) = \frac{1}{2}\epsilon(t,\hat{\theta}^{(j)}_{t-1|t-1})\epsilon^{\mathrm{T}}(t,\hat{\theta}^{(j)}_{t-1|t-1}). \tag{12}$$

The following modified artificial evolution law is now proposed for the parameter update in the particle filter for generating $j = 1, ..., M$ parameter particles that correspondingly determines the distribution from which the *a priori* parameter estimate $\hat{\theta}^{(j)}_{t|t-1}$ is obtained as follows,

$$\tilde{\theta}^{(j)}_{t|t-1} = \hat{\theta}^{(j)}_{t-1|t-1} + \gamma_t R^{(j)}_t \psi^{(j)}_t \epsilon(t,\hat{\theta}^{(j)}_{t-1|t-1}) + \zeta^{(j)}_t, \tag{13a}$$

$$m^{(j)}_t = \hat{\theta}^{(j)}_{t-1|t-1} + \gamma_t R^{(j)}_t \psi^{(j)}_t \epsilon(t,\hat{\theta}^{(j)}_{t-1|t-1}), \tag{13b}$$

$$\hat{\theta}^{(j)}_{t|t-1} = A_t m^{(j)}_t + (I - A_t)\bar{m}_{t-1} + \zeta^{(j)}_t, \tag{13c}$$

$$\bar{m}_{t-1} = \frac{1}{M}\sum_{j=1}^{M}\hat{\theta}^{(j)}_{t-1|t-1},$$

$$\bar{y}^{(j)}_{t|t-1} = h_t(\hat{x}_{t|t},\hat{\theta}^{(j)}_{t|t-1}), \tag{13d}$$

where $\tilde{\theta}^{(j)}_{t|t-1}$ denotes the particles obtained from the modified RPE-based artificial evolution law (13a), $\psi_t = \frac{\partial \hat{y}_t}{\partial \hat{\theta}_{t-1|t-1}} = \frac{\partial h_t(\hat{x}_{t|t},\hat{\theta}_{t-1|t-1})}{\partial \hat{\theta}_{t-1|t-1}}$, which when evaluated at $\hat{\theta}^{(j)}_{t-1|t-1}$ is denoted by $\psi^{(j)}_t$, γ_t is a time-varying step size design parameter, $\zeta^{(j)}_t \sim \mathcal{N}(0, (I - A^2_t)V_{\hat{\theta}_{t-1}})$ denotes the zero-mean normal increment particles to the parameter update law at each time step with the covariance matrix $(I - A^2_t)V_{\hat{\theta}_{t-1}}$ through the use of the kernel smoothing concept, A_t denotes the shrinkage matrix and $V_{\hat{\theta}_{t-1}}$ denotes the covariance of the parameter estimation in the previous time step $t-1$. The kernel shrinkage algorithm attempts to force the distribution of the parameters particles towards the mean of their distribution in the previous time instant as denoted by $\bar{m}_{t-1}$ by applying the shrinkage coefficient matrix A_t to the obtained $m^{(j)}_t$. The processes $\hat{\theta}^{(j)}_{t-1|t-1}$ and $\zeta^{(j)}_t$ are conditionally independent given observations up to time t. Moreover, $R^{(j)}_t = \mathbb{E}[\epsilon^{(j)}_t \epsilon^{(j)\mathrm{T}}_t]$ denotes a time-varying coefficient which determines the updating direction and is selected as a positive scalar to ensure that the criterion (12) can be minimized by changing $\tilde{\theta}^{(j)}_{t|t-1}$ in the steepest descent direction. The *a priori* parameter estimation particle is denoted by $\hat{\theta}^{(j)}_{t|t-1}$. The measurement equation (13d), is used to distinguish the evaluated output $\bar{y}^{(j)}_{t|t-1}$ obtained by the parameter estimation filter from the one obtained by the state estimation filter, as provided in the Subsection III-B. The equations (13a)-(13d) are to be used for implementing the parameter estimation particle filter scheme.

In Theorem 1 that is provided in the Subsection III-E, it will be shown that the update law (13a) can guarantee the convergence of the parameter estimate particles $\hat{\theta}^{(j)}_{t-1|t-1}$, $j = 1, ..., M$

(after the resampling step), to a local minimum of $\mathbb{E}(\bar{J}(\hat{\theta}^{(j)}_{t-1|t-1})|y_{1:t-1}, \hat{x}_t)$ which is located in a compact set of $\{x_t, \theta_t\}$ for which the functions $f_t(x_t, \theta_t, \omega_t)$, and $h_t(x_t, \theta_t)$ are sufficiently smooth with respect to x_t and θ_t, so that a small change in x_t or θ_t cannot result in a big change in their estimates. This set is denoted by $D_{\mathcal{M}}$.

The parameter update law (13a) is known as the RPE-based modified artificial evolution law. This update law contains a term in addition to the independent normal increment $\zeta^{(j)}_t$. The estimated parameter from this update law is invoked in the PF-based parameter estimation filter to represent the distribution from which the parameter particle population for the next time step is chosen. Therefore, the above proposed RPE-based modified artificial evolution law enables the PF-based estimation algorithm to handle and cope with the time-varying parameter scenarios. The time-varying term $\gamma_t R^{(j)}_t$, acts as an adaptive step size in equation (13a), therefore our algorithm can also be considered as an adaptive step size scheme.

In order to ensure that the obtained $\tilde{\theta}^{(j)}_{t|t-1}$ from the modified artificial evolution law given by equation (13a) remains in $D_{\mathcal{M}}$, the following projection algorithm is utilized that forces $\tilde{\theta}^{(j)}_{t|t-1}$ to remain inside the compact subset $\bar{D}$ of $D_{\mathcal{M}}$ ($\bar{D} \subset D_{\mathcal{M}}$) according to the following procedure [19],

1) Choose a factor $0 \leq \mu \leq 1$,
2) Compute $\breve{\theta}^{(j)}_{t|t-1} := \gamma_t R^{(j)}_t \psi^{(j)}_t (y_t - h_t(\hat{x}_{t|t}, \hat{\theta}^{(j)}_{t-1|t-1}))$,
3) Construct $\tilde{\theta}^{(j)}_{t|t-1} := \hat{\theta}^{(j)}_{t-1|t-1} + \breve{\theta}^{(j)}_{t|t-1}$,
4) If $\tilde{\theta}^{(j)}_{t|t-1} \in D_{\mathcal{M}}$ go to Step 6 else go to Step 5,
5) Set $\breve{\theta}^{(j)}_{t|t-1} = \mu \breve{\theta}^{(j)}_{t|t-1}$ and go to Step 3,
6) Stop.

Consequently, the one-step ahead prediction distribution of the parameter θ_t is now obtained as,

$$\tilde{\pi}^M_{\theta_{t|t-1}}(\mathrm{d}\theta_t) \triangleq \frac{1}{M} \sum_{j=1}^{M} \delta_{\hat{\theta}^{(j)}_{t|t-1}}(\mathrm{d}\theta_t). \tag{14}$$

In the measurement update step, the approximation to $\pi_{\theta_{t|t}}(\mathrm{d}\theta_t)$ that is denoted by $\tilde{\pi}^M_{\theta_{t|t}}(\mathrm{d}\theta_t)$, is given according to the Bayes' rule as $\tilde{\pi}^M_{\theta_{t|t}}(\mathrm{d}\theta_t) \propto \frac{\rho(y_t|x_t,\theta_t)\tilde{\pi}^M_{\theta_{t|t-1}}(\mathrm{d}\theta_t)\pi^N_{x_{t|t}}(\mathrm{d}x_t)}{\int_{\mathbb{R}^{n_\theta}}\int_{\mathbb{R}^{n_x}} \rho(y_t|x_t,\theta_t)\tilde{\pi}^M_{\theta_{t|t-1}}(\mathrm{d}\theta_t)\pi^N_{x_{t|t}}(\mathrm{d}x_t)}$. Since the state estimate at time t, i.e. $\hat{x}_{t|t}$ is assumed to be known and approximated as the mean of the *a posteriori* state estimate distribution $\pi^N_{x_{t|t}}(\mathrm{d}x_t)$ (from a filter with N particles), therefore, $\tilde{\pi}^M_{\theta_{t|t}}(\mathrm{d}\theta_t) \propto \frac{\rho(y_t|\hat{x}_{t|t},\hat{\theta}^{(j)}_{t|t-1})\delta_{\hat{\theta}^{(j)}_{t|t-1}}(\mathrm{d}\theta_t)}{\frac{1}{M}\sum_{j=1}^{M}\rho(y_t|\hat{x}_{t|t},\hat{\theta}^{(j)}_{t|t-1})}$.

Remark 2. *In order for the optimal parameter filter to exist, in the approximation of* $\pi_{\theta_{t|t}}(\mathrm{d}\theta_t)$ *one must ensure that* $\frac{1}{M}\sum_{j=1}^{M}\rho(y_t|\hat{x}_{t|t},\hat{\theta}_{t|t-1}^{(j)}) \geq r_{th}^2$, *where* r_{th}^2 *denotes a threshold that is selected for sufficiently large* M.

Consequently, as the present observation y_t becomes available, the particle weights $w_{\theta_t}^{(j)}$ are updated by the likelihood function according to $w_{\theta_t}^{(j)} \sim p_{\nu_t}(y_t - \bar{y}_{t|t-1}^{(j)}) = \rho(y_t|\hat{x}_{t|t},\hat{\theta}_{t|t-1}^{(j)})$, where $p_{\nu_t}(.)$ denotes the pdf of the measurement noise. This can now be expressed by using the normalized weights $\tilde{w}_{\theta_t}^{(j)}$ as $\tilde{\pi}_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t) = \sum_{j=1}^{M}\tilde{w}_{\theta_t}^{(j)}\delta_{\hat{\theta}_{t|t-1}^{(j)}}(\mathrm{d}\theta_t)$, where $\tilde{w}_{\theta_t}^{(j)} \triangleq \frac{\rho(y_t|\hat{x}_{t|t},\hat{\theta}_{t|t-1}^{(j)})}{\sum_{j=1}^{M}\rho(y_t|\hat{x}_{t|t},\hat{\theta}_{t|t-1}^{(j)})}$. Following the resampling/selection step, an equally weighted particle distribution $\pi_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t)$ is obtained as $\pi_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t) = \frac{1}{M}\sum_{j=1}^{M}\delta_{\bar{\theta}_{t|t-1}^{(j)}}(\mathrm{d}\theta_t)$ for approximating $\pi_{\theta_{t|t}}(\mathrm{d}\theta_t)$, and the resampled (selected) particles that are denoted by $\bar{\theta}_{t|t-1}^{(j)}$ follow the distribution $\tilde{\pi}_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t)$. It must be noted that in our proposed parameter estimation filter the state is taken as the available estimated $\hat{x}_{t|t}$. Therefore, the *a posteriori* parameter estimation distribution is approximated by a weighted sum of Dirac-delta masses as $\tilde{\pi}_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t)$ before performing the resampling and with an equally weighted particle distribution approximation as $\pi_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t)$ according to

$$\begin{aligned}
&\tilde{\pi}_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t) \approx \sum_{j=1}^{M}\tilde{w}_{\theta_t}^{(j)}\delta_{\hat{\theta}_{t|t-1}^{(j)}}(\mathrm{d}\theta_t),\\
&\tilde{w}_{\theta_t}^{(j)} \triangleq \frac{\rho(y_t|\hat{x}_{t|t},\hat{\theta}_{t|t-1}^{(j)})}{\sum_{j=1}^{M}\rho(y_t|\hat{x}_{t|t},\hat{\theta}_{t|t-1}^{(j)})},\\
&\pi_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t) = \frac{1}{M}\sum_{j=1}^{M}\delta_{\bar{\theta}_{t|t-1}^{(j)}}(\mathrm{d}\theta_t) \rightarrow \hat{\theta}_{t|t} = \frac{1}{M}\sum_{j=1}^{M}\bar{\theta}_{t|t-1}^{(j)}
\end{aligned} \tag{15}$$

where $\tilde{w}_{\theta_t}^{(j)}$ denotes the normalized parameter particle weight, $\{\bar{\theta}_{t|t-1}^{(j)}\}_{j=1}^{M}$ is obtained from the resampling/selection step of the scheme by duplicating the particles $\hat{\theta}_{t|t-1}^{(j)}$ having large weights and discarding the ones with small values to emphasize on the zones with higher *a posteriori* probabilities according to $P(\bar{\theta}_{t|t-1}^{(j)} = \hat{\theta}_{t|t-1}^{(k)}) = \tilde{w}_{\theta_t}^{(k)},\ k = 1,...,M$. In our proposed filter the residual resampling method is used to ensure that the variance reduction among the resampled particles is guaranteed [33].

Therefore, an approximation to $\mathbb{E}(\phi(\theta_t)|y_{1:t},x_t)$ by $\phi(\theta_t) = \theta_t$ takes on the form $\hat{\theta}_{t|t} \sim \pi_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t) = \frac{1}{M}\sum_{j=1}^{M}\delta_{\bar{\theta}_{t|t-1}^{(j)}}(\mathrm{d}\theta_t)$, where $\pi_{\theta_{t|t}}^{M}(\mathrm{d}\theta_t)$ denotes the *a posteriori* distribution of the parameter estimate (after performing the resampling from $\hat{\theta}_{t|t-1}^{(j)}$). The resulting estimated output of this filter is obtained by $\hat{y}_t = h_t(\hat{x}_{t|t},\hat{\theta}_{t|t})$. The explicit details for implementation of the

parameter estimation filter are now provided below.

<u>*The Parameter Estimation Filter*</u>

The particle filter for implementation of the parameter estimation is described as follows:

1) Initialize the M particles for the parameters as $\{\theta_0^j\}_{j=1}^M \sim \pi_{\theta_0}(\mathrm{d}\theta_0)$, and use the initial values of the states as x_0 which represents the mean of the states initial distribution $\pi_{x_0}(\mathrm{d}x_0)$.
2) Draw $\zeta_t^{(j)} \sim \mathcal{N}(0, (I - A_t^2)V_{\hat{\theta}_{t-1|t-1}})$.
3) Predict $\hat{\theta}_{t|t-1}^{(j)},\ j = 1, ..., M$ from equations (13a)-(13c).
4) If $\frac{1}{M}\sum_{j=1}^M \rho(y_t|\hat{x}_{t|t}, \hat{\theta}_{t|t-1}^{(j)}) \geq r_{th}^2$ (according to Remark 2) go to Step 5, otherwise return to Step 2.
5) Compute the importance weights $\{w_{\theta_t}^{(j)}\}_{j=1}^M, w_{\theta_t}^{(j)} = \rho(y_t|\hat{x}_{t|t}, \hat{\theta}_{t|t-1}^{(j)}), j = 1, ..., M$, and normalize them to $\tilde{w}_{\theta_t}^{(j)} = \frac{w_{\theta_t}^{(j)}}{\sum_{j=1}^M w_{\theta_t}^{(j)}}$.
6) Resampling: Draw M new particles with replacement for each $j = 1, ..., M$, $P(\bar{\theta}_{t|t-1}^{(j)} = \hat{\theta}_{t|t-1}^{(k)}) = \tilde{w}_{\theta_t}^{(k)},\ k = 1, ..., M$, where $\hat{\theta}_{t|t-1}^{(j)} \sim \tilde{\pi}_{\theta_{t|t}}^M(\mathrm{d}\theta_t) = \sum_{j=1}^M \tilde{w}_{\theta_t}^{(j)} \delta_{\hat{\theta}_{t|t-1}^{(j)}}(\mathrm{d}\theta_t)$.
7) Construct $\hat{\theta}_{t|t}$ from the conditional distribution $\pi_{\theta_{t|t}}^M(\mathrm{d}\theta_t) = \frac{1}{M}\sum_{j=1}^M \delta_{\bar{\theta}_{t|t-1}^{(j)}}(\mathrm{d}\theta_t)$ with equally weighted $\bar{\theta}_{t|t-1}^{(j)}$ as $\hat{\theta}_{t|t} = \frac{1}{M}\sum_{j=1}^M \bar{\theta}_{t|t-1}^{(j)}$.
8) Set $t = t + 1$ and go to Step 2 of the state estimation filter as provided in the Subsection III-B.

As stated earlier the kernel from which the parameter particles i.e. $\hat{\theta}_{t|t-1}^{(j)}$ for the next time step is chosen is a Gaussian kernel and its mean is obtained from $m_t^{(j)}$ and its variance is obtained based on the kernel smoothing consideration that is provided in the next Subsection III-D. In the subsections below the required conditions for the boundedness of the parameter transition kernel $K_\theta(.)$ are also investigated and developed.

D. Kernel Smoothing of the Parameters

In this subsection, the kernel smoothing approach [18] is utilized to ensure that the variance of the normal distribution which is obtained based on the modified artificial evolution law for the parameter estimates remains bounded.

Consider the modified artificial evolution law (13a) in which $\zeta_t^{(j)}$ is a normal zero-mean uncorrelated random increment to the parameter that is estimated at time $t-1$. As $t \to \infty$, the

variance of the added evolution increases and can therefore make $\tilde{\theta}^{(j)}_{t|t-1}$ in the modified RPE-based artificial evolution (13a), and consequently $\hat{\theta}^{(j)}_{t|t-1}$ in (13c) become completely unreliable. This phenomenon is known as the loss of information that can also occur between two consequent sampling times. On the other hand, since $\hat{\theta}_{t|t}$ is time-varying, generally there will not exist an optimal value for the variance of the evolution noise $\zeta^{(j)}_t$ that remains suitable for all times.

Consequently, the idea of the kernel shrinkage has been proposed in [18]. In the kernel shrinkage approach [17], for the next time step one takes the mean of the estimated parameter distribution in the particle filter according to

$$\begin{aligned} K_\theta(\mathrm{d}\theta_t|\theta^{(j)}_{t-1}, x_t) &\approx p(\hat{\theta}^{(j)}_{t|t}|\hat{\theta}^{(j)}_{t-1|t-1}) \\ &= \mathcal{N}(A_t m^{(j)}_t + (I - A_t)\bar{m}_{t-1}, (I - A_t^2)V_{\hat{\theta}_{t-1|t-1}}), \end{aligned} \tag{16}$$

where $m^{(j)}_t$ for $j = 1, ..., M$, is obtained from (13b). By utilizing this kernel shrinkage rule, the resulting normal distribution retains the mean $\bar{m}_{t-1}$ and has the appropriate variance for avoiding over-dispersion relative to the *a posteriori* sample. The kernel shrinkage forces the parameter samples towards their mean before the noise $\zeta^{(j)}_t$ is added. In our proposed approach the changes due to the parameter variations are considered in the mean of the parameter estimate distribution through the modified artificial evolution rule, therefore the mean of the distribution, i.e. $\bar{m}_{t-1}$, itself is time-varying and the kernel shrinkage ensures a smooth transition in the estimated parameters even when they are subjected to changes. To eliminate the information loss effect, the variance taken from both sides of equation (13c) results in $V_{\hat{\theta}_{t|t-1}} = A_t^2 V_{\hat{\theta}_{t-1|t-1}} + (I - A_t^2)V_{\hat{\theta}_{t-1|t-1}} = V_{\hat{\theta}_{t-1|t-1}}$, that ensures the variance of the added random evolution would not cause over-dispersion in the parameter estimation algorithm for all time.

The following proposition specifies an upper bound on the shrinkage factor. This upper bound is calculated in the worst case, that is when the parameter is considered to be constant but the modified evolution law (13a) is used in the parameter estimation filter for estimating it. Utilization of this upper bound in the kernel shrinkage algorithm ensures the boundedness of the variance of the estimated parameters distribution that is obtained according to the RPE-based artificial evolution update law and the kernel smoothing augmented with the shrinkage factor.

Proposition 1: *Upper bound on the kernel shrinkage factor*: Given the parameter update law (13a), the estimated parameters conditional normal distribution based on the kernel smoothing as given by equation (16), result in an upper bound for A_t that is obtained as $A_t \leq I(1 -$

$[\frac{1}{2}\frac{\sigma_{\min}(W_t V^{-1}_{\hat{\theta}_{t-1|t-1}} + P^2_{t_{\max}}\Psi V_y \Psi^{\mathrm{T}} V^{-1}_{\hat{\theta}_{t-1|t-1}})}{\sigma_{\max}(W_t V^{-1}_{\hat{\theta}_{t-1|t-1}} + P^2_{t_{\max}}\Psi V_y \Psi^{\mathrm{T}} V^{-1}_{\hat{\theta}_{t-1|t-1}})}])$, where $\psi_t^{(j)}$ in equation (13a) is considered as constant between the time steps t and $t-1$ and is denoted by Ψ. Moreover, $\sigma_{\min}$ and $\sigma_{\max}$ denote the minimum and the maximum eigenvalues of a matrix, respectively, W_t denotes the variance of the added noise, V_y denotes the variance of the measurement noise, and $P_{t_{\max}} = \max\{\gamma_t R_t^{(j)}\}_{j=1}^{M}$.

Proof: Let us consider the modified artificial evolution equation (13a). Let $V(.|y_{1:t})$ denote the variance of a given stochastic process assuming that the observations up to time t are available, and $C(.,.|y_{1:t})$ denote the covariance of two given stochastic processes by assuming that the observations up to time t are available. In our method the variations in the time-varying parameters are captured through the prediction error term $\epsilon_t^{(j)}$ and then used in the modified artificial evolution model (13a) in order to determine the mean of the normal distribution from which the parameter population for the next time step is chosen. Our main concern here relates to the case when the actual system parameter is constant but the RPE term $P_t^{(j)}\psi_t^{(j)}\epsilon_t^{(j)}$, and the evolution noise $\zeta^{(j)}$ are still added to the parameter estimate rule. In this case, it is necessary to introduce correlations between the predicted parameter $\tilde{\theta}^{(j)}_{t|t-1}$ and the random noise $\zeta_t^{(j)}$.

By taking into account the relationship between the variance of both sides of equation (13a) when the covariance matrix is assumed to be non-singular, and by assuming that $\{\psi_t^{(j)}\}_{i=1}^{M} = \Psi$ is constant between $t-1$ and t, yields $V(\tilde{\theta}^{(j)}_{t|t-1}|y_{1:t}) = V(\hat{\theta}^{(j)}_{t-1|t-1}|y_{1:t}) + P_t^{(j)^2}\Psi V_y \Psi^{\mathrm{T}} + W_t + 2C(\hat{\theta}^{(j)}_{t-1|t-1}, \zeta_t^{(j)}|y_{1:t}) + 2C(P_t^{(j)}\Psi\epsilon_t^{(j)}, \zeta_t^{(j)}|y_{1:t}) + 2C(\hat{\theta}^{(j)}_{t-1|t-1}, P_t^{(j)}\Psi\epsilon_t^{(j)}|y_{1:t})$. Furthermore, since $P_t^{(j)} = \gamma_t R_t^{(j)} = \gamma_t \mathbb{E}(\epsilon_t^{(j)}\epsilon_t^{(j)\mathrm{T}})$ is a scalar, one can write $V(P_t^{(j)}\Psi\epsilon_t^{(j)}|y_{1:t}) = (\mathbb{E}[P_t^{(j)}|y_{1:t}])^2 V(\Psi\epsilon_t^{(j)} |y_{1:t}) + (\mathbb{E}[\Psi\epsilon_t^{(j)}|y_{1:t}])^2 V(P_t^{(j)}|y_{1:t}) + V(P_t^{(j)}|y_{1:t})V(\Psi\epsilon_t^{(j)}|y_{1:t}) = P_t^{(j)^2}\Psi V_y \Psi^{\mathrm{T}}$. In order to ensure that there is no information loss due to the modified artificial evolution law, one must have, $V(\tilde{\theta}_{t|t-1}|y_{1:t}) = V(\hat{\theta}_{t-1|t-1}|y_{1:t}) = V_{\hat{\theta}_{t-1|t-1}}$, which implies that, $C(\hat{\theta}_{t-1}, \zeta_t|Y_t) + C(P_t\Psi\epsilon_t, \zeta_t|Y_t) = -\frac{1}{2}W_t - \frac{1}{2}P_t^2\Psi V_y \Psi^{\mathrm{T}}$, where $C(\hat{\theta}^{(j)}_{t-1|t-1}, P_t^{(j)}\Psi\epsilon_t^{(j)}|y_{1:t})$ is ignored since it is assumed that when the parameter is constant, it is then independent from $\epsilon_t^{(j)}$.

Therefore, negative correlations are needed to remove the effects of the unwanted information loss. In case of approximate joint normality of $(\hat{\theta}^{(j)}_{t-1|t-1}, \zeta_t^{(j)}|Y_t)$ and $(P_t^{(j)}\Psi\epsilon_t^{(j)}, \zeta_t^{(j)}|Y_t)$, the conditional normal evolution is obtained as $p(\hat{\theta}^{(j)}_{t|t}|\hat{\theta}^{(j)}_{t-1|t-1}) \sim \mathcal{N}(A_t m_t^{(j)} + (I - A_t)\bar{m}_{t-1}, (I - A_t^2)V_{\hat{\theta}_{t-1|t-1}})$, where $m_t^{(j)}$ at each time step is found from equation (13b). The shrinkage matrix A_t, is obtained as $A_t = I - [\frac{1}{2}(W_t V^{-1}_{\theta_{t-1|t-1}} + P^2_{t_{\max}}\Psi V_y \Psi^{\mathrm{T}} V^{-1}_{\theta_{t-1|t-1}})]$. The upper bound for A_t is obtained as $A_t \leq I(1 - [\frac{1}{2}\frac{\sigma_{\min}(W_t V^{-1}_{\hat{\theta}_{t-1|t-1}} + P^2_{t_{\max}}\Psi V_y \Psi^{\mathrm{T}} V^{-1}_{\hat{\theta}_{t-1|t-1}})}{\sigma_{\max}(W_t V^{-1}_{\hat{\theta}_{t-1|t-1}} + P^2_{t_{\max}}\Psi V_y \Psi^{\mathrm{T}} V^{-1}_{\hat{\theta}_{t-1|t-1}})}])$, where $P_{t_{\max}}$ is the upper bound

of $P_t^{(j)}$, $j = 1, ..., M$, and the normalization of the eigenvalue is performed to ensure that the related fraction remains less than 1. Let $a = 1 - [\frac{1}{2}\frac{\sigma_{\min}(W_t V^{-1}_{\hat{\theta}_{t-1|t-1}} + P^2_{t_{\max}} \Psi V_y \Psi^{\mathrm{T}} V^{-1}_{\hat{\theta}_{t-1|t-1}})}{\sigma_{\max}(W_t V^{-1}_{\hat{\theta}_{t-1|t-1}} + P^2_{t_{\max}} \Psi V_y \Psi^{\mathrm{T}} V^{-1}_{\hat{\theta}_{t-1|t-1}})}]$, therefore, the shrinkage matrix becomes $A_t = aI$ with the shrinkage factor a. The smoothing matrix of the normal distribution variance is now obtained from the shrinkage factor as $(1 - a^2)I$, which guarantees that the normal distribution of the parameter estimates has a finite variance for all time. This completes the proof of the proposition. ■

The convergence of the estimated parameter particles $\hat{\theta}^{(j)}_{t-1|t-1}$, $j = 1, ..., M$ to the local minimum of $\mathbb{E}(\bar{J}(\hat{\theta}^{(j)}_{t-1|t-1})|y_{1:t-1}, \hat{x}_t)$ is now investigated in the following subsection. The developed convergence proof does not ensure the convergence of the RPE-based parameter estimation method to the true parameter value, but only to a set of zeros of the gradient of the chosen cost function.

E. Convergence of the RPE-based Parameter Update Law

In order to investigate the convergence of our proposed RPE-based modified artificial evolution law for updating the parameter particles distribution and to achieve a local minimization of $\mathbb{E}(\bar{J}(\theta_{t-1})|y_{1:t-1}, x_t)$, consider equation (13a), where γ_t denotes a time-varying step size such that $\lim_{t\to\infty} \gamma_t = \mu_0 > 0$, where μ_0 is a small positive constant. The introduction of the step size γ_t is necessary to transform the discrete-time model (13a) into a continuous-time representation as shown subsequently.

First, we need to state the following three assumptions A1-A3 according to [19], to guarantee the convergence of our proposed algorithm as presented in our main result below in Theorem 1. Specifically, let us assume:

A1. The vector $\{x_t, \theta_t\}$ is defined such that it ranges over a compact set denoted by $D_{\mathcal{M}}$, for which the functions $f_t(x_t, \theta_t, \omega_t)$ and $h_t(x_t, \theta_t)$ are continuously differentiable with respect to the state x_t as well as the parameter θ_t.

A2. The function $l(\epsilon(t, \hat{\theta}_{t-1|t-1}))$ is sufficiently smooth and twice continuously differentiable w.r.t. ϵ, and $|l_{\epsilon\epsilon}(\epsilon(t, \hat{\theta}_{t-1|t-1}))| \leq C$ for $\hat{\theta}_{t-1|t-1} \in D_{\mathcal{M}}$, where $l_{\epsilon\epsilon}(\epsilon(t, \hat{\theta}_{t-1|t-1}))$ denotes the second derivative of $l(\epsilon(t, \hat{\theta}_{t-1|t-1}))$ w.r.t. ϵ.

A3. The data generation sequence y_t (generated from equation (2)), is such that $\bar{\mathbb{E}}(l(\epsilon(t, \hat{\theta}_{t-1|t-1})))$ $= \bar{J}(\hat{\theta}_{t-1|t-1})$ and $\bar{\mathbb{E}}[\frac{\mathrm{d}}{\mathrm{d}\hat{\theta}_{t-1|t-1}} l(\epsilon(t, \hat{\theta}_{t-1|t-1}))] = -g(\hat{\theta}_{t-1|t-1})$ exist for all $\hat{\theta}_{t-1|t-1} \in D_{\mathcal{M}}$, where $\bar{\mathbb{E}}(l(\epsilon(t, \hat{\theta}_{t-1|t-1}))) = \frac{1}{K} \sum_{\tau=t-\kappa}^{t} \mathbb{E} l(\epsilon(\tau, \hat{\theta}_{t-1|t-1}))$.

It must be noted that the kernel shrinkage method, as stated earlier, attempts to retain the mean of the parameter estimation particles at time t near the estimated parameter in the previous time step $t-1$, i.e. $\hat{\theta}_{t-1|t-1}$. Therefore, in the following theorem the convergence properties of $\hat{\theta}_{t-1|t-1}$ is now addressed. The main result of this section is stated below.

Theorem 1. *Consider the parameter estimation algorithm as specified by the equations* (13a)-(13d). *Also consider the a posteriori parameter estimate governed by equation* (15). *Let Assumptions A1 to A3 hold. It now follows that the particles* $\hat{\theta}^{(j)}_{t-1|t-1}$, $j = 1, ..., M$, *and consequently the distribution of the estimated parameter particles approximated by the particle filter* $\pi^M_{\theta_{t-1|t-1}}(\mathrm{d}\theta_{t-1})$, *w.p.1 converge either to the set* $D_{\mathcal{C}} = \{\hat{\theta}^{(i)}_{t-1|t-1} | \hat{\theta}^{(i)}_{t-1|t-1} \in D_{\mathcal{M}},\ \frac{d}{d\hat{\theta}^{(j)}_{t-1|t-1}} \bar{J}(\hat{\theta}^{(j)}_{t-1|t-1}) = 0,\ j = 1, ..., M\}$ *or to the boundary of* $D_{\mathcal{M}}$ *as* $t \to \infty$.

Proof: The existence of the projection algorithm in the parameter estimation scheme ensures that $\tilde{\theta}^{(j)}_{t|t-1}$ remains inside $D_{\mathcal{M}}$. According to equation (15), the *a posteriori* estimate of the parameter at time t is obtained from the resampled particles of the *a priori* parameter estimate $\bar{\theta}^{(j)}_{t|t-1}$, as $\hat{\theta}_{t|t} = \frac{1}{M}\sum_{j=1}^{M} \bar{\theta}^{(j)}_{t|t-1}$, where $\bar{\theta}^{(j)}_{t|t-1}$, is selected from the M particles of $\hat{\theta}^{(j)}_{t|t-1}$ for which $\rho_{\nu_t}(y_t - h(\hat{x}_{t|t}, \hat{\theta}^{(j)}_{t|t-1}))$ has higher probabilities.

Consider equation (13c) for generating $\hat{\theta}^{(j)}_{t|t-1}$ and substitute $m^{(j)}_t$ from the RPE-based update rule of equations (13a)-(13b), to obtain the following expression for the resampled particles $\bar{\theta}^{(j)}_{t|t-1}$, namely

$$\bar{\theta}^{(j)}_{t|t-1} = A_t \hat{\theta}^{(j)}_{t-1|t-1} + (I - A_t)\frac{1}{M}\sum_{j=1}^{M} \hat{\theta}^{(j)}_{t-1|t-1} + A_t \gamma_t R^{(j)}_t \psi^{(j)}_t \epsilon(t, \hat{\theta}^{(j)}_{t-1|t-1}) + \sqrt{I - A_t^2}\,\zeta^{(j)}_t, \tag{17}$$

By applying the sum operator to both sides of equation (17) to construct $\hat{\theta}_{t|t}$ yields,

$$\frac{1}{M}\sum_{j=1}^{M} \bar{\theta}^{(j)}_{t|t-1} = A_t \frac{1}{M}\sum_{j=1}^{M} \hat{\theta}^{(j)}_{t-1|t-1} + \frac{1}{M}\sum_{j=1}^{M}\frac{1}{M}\sum_{j=1}^{M} \hat{\theta}^{(j)}_{t-1|t-1} - A_t \frac{1}{M}\sum_{j=1}^{M}\frac{1}{M}\sum_{j=1}^{M} \hat{\theta}^{(j)}_{t-1|t-1}$$

$$+ A_t \frac{1}{M}\sum_{j=1}^{M} \gamma_t R^{(j)}_t \psi^{(j)}_t \epsilon(t, \hat{\theta}^{(j)}_{t-1|t-1}) + \sqrt{I - A_t^2}\frac{1}{M}\sum_{j=1}^{M} \zeta^{(j)}_t = \frac{1}{M}\sum_{j=1}^{M} \hat{\theta}^{(j)}_{t-1|t-1} + A_t \frac{1}{M}\sum_{j=1}^{M} \gamma_t R^{(j)}_t \psi^{(j)}_t \epsilon(t, \hat{\theta}^{(j)}_{t-1|t-1}),$$

which results in $\hat{\theta}_{t|t} = \hat{\theta}_{t-1|t-1} + A_t \frac{1}{M}\sum_{j=1}^{M} \gamma_t R^{(j)}_t \psi^{(j)}_t \epsilon(t, \hat{\theta}^{(j)}_{t-1|t-1})$. Assumptions A1 and A2 ensure that the regularity conditions are satisfied according to [19], consequently, a differential equation associated with equation (13a) can be obtained by considering that $\Delta\tau$ is a sufficiently small number and $t, \check{t}$ are specified such that $\sum_{k=t}^{\check{t}} \gamma_k = \Delta\tau$. Through a change of time-scales as $t \to \tau$ and $\check{t} \to \tau + \Delta\tau$ for a sufficiently small $\Delta\tau$, and by assuming that $\hat{\theta}_{t-1|t-1} = \check{\theta}, R^{(j)}_t = \check{R}^{(j)},\ A_t = aI$ is a constant matrix, the difference equation for $\hat{\theta}^{(j)}_{t|t}$ is now expressed as

$$\hat{\theta}^{(j)}_{\check{t}} \approx \check{\theta}^{(j)} + a\Delta\tau \check{R}^{(j)} g(\check{\theta}^{(j)}), \tag{18}$$

where $g(\check{\theta}^{(j)}) = \frac{1}{\Delta\tau}\sum_{k=t}^{t=\check{t}} \psi_t^{(j)} \epsilon(k, \hat{\theta}_{t-1|t-1}^{(j)})$. Consequently, considering Assumption A3, the differential equation related to the evolution of each single particle is obtained as,

$$\frac{\mathrm{d}\theta_D^{(j)}}{\mathrm{d}\tau} = aR_D^{(j)}(\tau)g(\hat{\theta}_D^{(j)}(\tau)) = -aR_D^{(j)}(\tau)[\frac{d}{d\hat{\theta}_D^{(j)}}\bar{J}(\hat{\theta}_D^{(j)})]^{\mathrm{T}}, \tag{19}$$

where the subscript D is used to differentiate the solution of the differential equation (19) from the solution of the difference equation (18). Now, the required convergence analysis is reduced to showing the properties of the deterministic continuous-time system (19).

Consider the function $L(\hat{\theta}_{t-1|t-1}^{(j)}) = \mathbb{E}(\bar{J}(\hat{\theta}_{t-1|t-1}^{(j)}) = \frac{1}{M}\sum_{j=1}^{M} \bar{J}(\hat{\theta}_{t-1|t-1}^{(j)})$ representing the expectation of a positive definite function through M data points for $\hat{\theta}_{t-1|t-1}^{(j)}$, which is itself a positive definite function. Our goal is to evaluate the derivative of this function along the trajectories of the system (19). According to Assumption A2, the second derivative of $l(\epsilon(t, \hat{\theta}_{t-1|t-1}^{(j)}))$ is bounded, therefore the summation and derivative operations commute. According to Assumption A3 for $\hat{\theta}_{t-1|t-1}^{(j)} \in D_{\mathcal{M}}$, $\acute{\bar{J}}(\hat{\theta}_D^{(j)}(\tau)) = \frac{d}{d\hat{\theta}_{t-1|t-1}^{(j)}}\bar{J}(\hat{\theta}_{t-1|t-1}^{(j)})|_{\hat{\theta}_{t-1|t-1}^{(j)}=\check{\theta}^{(j)}} = \bar{\mathbb{E}}(\frac{d}{d\hat{\theta}_{t-1|t-1}}l(\epsilon(t, \hat{\theta}_{t-1|t-1}^{(j)}))$, exists and is approximated by $-g(\check{\theta}^{(j)})$. Therefore, considering that $a > 0$, and $R_D^{(j)}(\tau)$ is a positive scalar for $j = 1, ..., M$ (it represents the trace of a positive definite matrix at time τ), one gets

$$\frac{d}{d\tau}V(\hat{\theta}_D^{(j)}) = \mathbb{E}(\frac{d}{d\tau}\bar{J}(\hat{\theta}_D^{(j)}(\tau)) = \frac{1}{M}\sum_{j=1}^{M}\acute{\bar{J}}(\hat{\theta}_D^{(j)})\frac{d}{d\tau}\hat{\theta}_D^{(j)}(\tau) = \frac{-a}{M}\sum_{j=1}^{M}[g(\hat{\theta}_D^{(j)}(\tau))]R_D^{(j)}(\tau)[g(\hat{\theta}_D^{(j)})(\tau)]^{\mathrm{T}} \leq 0,$$

where the equality is obtained only for $\hat{\theta}_D(\tau) \in D_{\mathcal{C}}$. Therefore, as $t \to \infty$ either $\hat{\theta}_{t-1|t-1}^{(j)}$ and consequently, $\pi_{\theta_{t-1|t-1}}^{M}$ w.p.1 tend to $D_{\mathcal{C}}$ or to the boundary of $D_{\mathcal{M}}$, where w.p.1 is with respect to the random variable related to the parameter estimate particles. This completes the proof of the theorem. ■

It was indicated earlier that the above theorem can only guarantee the boundedness of the estimated parameter distribution from the particle filters and not its convergence to the optimal distribution. Based on the results of Theorem 1 and Proposition 1 one can ensure that the probability density function $p(\hat{\theta}_{t|t}^{(j)}|\hat{\theta}_{t-1|t-1}^{(j)})$ in the particle filter remains bounded, and therefore the related kernel $K_\theta(\mathrm{d}\theta_t|\theta_{t-1}, x_{t-1})$ (in the particle filter) is also bounded.

Remark 3. *Assuming that Remarks 1 and 2 are satisfied, and also by assuming that $\rho(y_s|x_s, \hat{\theta}_s) < \infty$, and $K_x(x_s|x_{s-1}, \hat{\theta}_{s-1}) < \infty$, the boundedness of the parameter estimation transition kernel $K_\theta(\hat{\theta}_s|\hat{\theta}_{s-1}, \hat{x}_s)$ is also ensured from the Proposition 1 and Theorem 1. Therefore, the convergence of the proposed dual state/parameter estimation filter to their optimal values, for $\{x_t, \theta_t\} \in D_{\mathcal{M}}$ can be guaranteed according to Theorem 3.1 in [31].*

Our proposed dual state and parameter estimation scheme is an effective scheme for the purpose of fault diagnosis of nonlinear systems where without loss of any generality one initiates operating the system from the healthy mode of operation. During the healthy operation of the system our proposed dual state and parameter estimation strategy can provide one with an accurate and reliable information on the health parameters of the system. This information can then be readily used to perform the tasks of fault detection, isolation and identification, following the presence of faults in the actuators, sensors, or the components of the system. In the following subsection, the application of our proposed approach for addressing the fault diagnosis of nonlinear systems is investigated.

F. The Fault Diagnosis Formulation

Determination and diagnosis of drifts in unmeasurable parameters of a system require an on-line parameter estimation scheme. In parametric modeling of a system anomaly or drift, generally it is assumed that the parameters are either constant or dependent on only the system states [34]. Hence, drifts in the parameters must be estimated through estimation techniques.

In [35], various parameter estimation methods such as least squares, instrumental variables and estimation via discrete-time models have been surveyed. The main drawbacks with such methods arise due to the complexity and nonlinearity of the system that make parameter estimation schemes a nonlinear optimization problem that must be solved in real-time. In [36] a nonlinear least squares (NLS) optimization scheme is developed for fault identification of a hybrid system.

The fault diagnosis problem under consideration in this work deals with obtaining an optimal estimate of the states as well as the time-varying parameters of a nonlinear system whose dynamics is governed by the discrete-time stochastic model,

$$x_{t+1} = f_t(x_t, \theta_t^{\mathrm{T}} \lambda(x_t), \omega_t), \tag{20}$$

$$y_t = h_t(x_t, \theta_t^{\mathrm{T}} \lambda(x_t)) + \nu_t, \tag{21}$$

where $f_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_\omega} \longrightarrow \mathbb{R}^{n_x}$ is a known nonlinear function, $\theta_t \in \mathbb{R}^{n_\theta}$ is the unknown and possibly time-varying parameter vector that for a healthy system is set to $\mathbf{1}$, $\lambda_t : \mathbb{R}^{n_x} \longrightarrow \mathbb{R}^{n_\theta}$ is a differentiable function that determines the relationship between the system states and the health parameters. The function $h_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \longrightarrow \mathbb{R}^{n_y}$ is a known nonlinear function, ω_t and ν_t are uncorrelated noise sequences with covariance matrices Q_t and R_t, respectively. According

to the formulation that is used in equations (20) and (21), the parameters θ_t to be estimated are considered as a multiplicative fault vector whose value is taken to be equal to $\mathbf{1}$ in the healthy mode of the system operation.

The model (20) and (21) is now used to investigate the problem of fault diagnosis (FD), which in this work implies and signifies the problem of fault detection, isolation, and identification (FDI) when the system health parameters are considered to be affected by a multiplicative fault vector θ_t. The system health parameters are known functions of the system states, $\lambda(x_t)$, and the multiplicative fault vector θ_t which is to be estimated. In other words, the estimated parameter $\hat{\theta}_{t|t}$ will be used to generate residual signals for accomplishing the fault diagnosis goal and objective. The required residuals are obtained as the difference between the estimated values of the parameters in the healthy operational mode and the faulty operational mode of the system, that is

$$r_t = \hat{\theta}_0 - \hat{\theta}_{t|t}, \tag{22}$$

where $\hat{\theta}_0$ denotes the estimated value of the parameter in the healthy operational mode and $\hat{\theta}_{t|t}$ denotes its *a posteriori* estimated value after the occurrence of the fault. It should be pointed out that the true value of the parameter is denoted by $\theta_t^\star$, which is clearly not available in reality.

In parameter estimation methods used for fault diagnosis of the system components, the residuals can be generated by comparing the estimated parameters that are obtained by either the ordinary least squares (OLS) or the recursive least squares (RLS) algorithms with the parameters that are obtained from estimation filters in the fault free case [37]. In the implementation of our proposed fault diagnosis strategy based on our developed state/parameter estimation framework, the parameter estimates will be considered as the main indicator to be used for detecting, isolating, and identifying the faults in the system components. The residuals are generated from the parameter estimates under the healthy and faulty operational modes of the system according to equation (22) as follows,

$$\hat{\theta}_0 = \text{argmax}(-\log(\hat{p}(\theta_0|y_{1:T})), \tag{23}$$

where $\hat{\theta}_0$ denotes the estimated parameters with the probability density $\hat{p}(\theta_0|y_{0:T})$ (conditioned on the observations up to time T), that are obtained from the collected estimates under the healthy operational mode of the system and fitted to a normal distribution. The time window T is chosen according to the convergence time of the parameter estimation algorithm. The thresholds

to indicate the confidence intervals for each parameter are obtained from Monte Carlo analysis that is performed under different single-fault and multi-fault scenarios. The estimated parameters $\hat{\theta}_{t|t}$ are generated by following the procedure that was developed and proposed in the previous sections of this work.

IV. Fault Diagnosis of a Gas Turbine Engine

In this section, the utility of our proposed dual estimation framework when applied to the problem of fault diagnosis of a nonlinear model of a gas turbine engine is demonstrated and investigated.

A. Model Overview

The utility and application of our proposed PF method for state/parameter estimation in a gas turbine engine is presented in the next subsections. The performance of our proposed state/parameter estimation scheme is evaluated and investigated subject to deficiencies in the system health parameters due to injection of simultaneous faults.

The mathematical model of a gas turbine that is used in this paper is a single spool jet engine as developed in [29]. The four engine states are the combustion chamber pressure and temperature, P_{CC} and T_{CC}, respectively, the spool speed N, and the nozzle outlet pressure P_{NLT}. The continuous-time state space model of the gas turbine is given as follows,

$$\begin{aligned}
\dot{T}_{\mathrm{CC}} &= \frac{1}{c_{\mathrm{v}}\dot{m}_{\mathrm{cc}}}[(c_{\mathrm{p}}T_{\mathrm{C}}\theta_{m_{\mathrm{C}}}\dot{m}_{\mathrm{C}} + \eta_{\mathrm{CC}}H_{\mathrm{u}}\dot{m}_{\mathrm{f}} - c_{\mathrm{p}}T_{\mathrm{CC}}\theta_{m_{\mathrm{T}}}\dot{m}_{\mathrm{T}}) - c_{\mathrm{v}}T_{\mathrm{CC}}(\theta_{m_{\mathrm{C}}}\dot{m}_{\mathrm{C}} + \dot{m}_{\mathrm{f}} - \theta_{m_{\mathrm{T}}}\dot{m}_{\mathrm{T}})],\\
\dot{N} &= \frac{\eta_{\mathrm{mech}}\theta_{m_{\mathrm{T}}}\dot{m}_{\mathrm{T}}c_{\mathrm{p}}(T_{\mathrm{CC}} - T_{\mathrm{T}}) - \theta_{m_{\mathrm{C}}}\dot{m}_{\mathrm{C}}c_{\mathrm{p}}(T_{\mathrm{C}} - T_{\mathrm{d}})}{\mathrm{J}N(\frac{\pi}{30})^2},\\
\dot{P}_{\mathrm{CC}} &= \frac{P_{\mathrm{CC}}}{T_{\mathrm{CC}}}\frac{1}{c_{\mathrm{v}}\dot{m}_{\mathrm{cc}}}[(c_{\mathrm{p}}T_{\mathrm{C}}\theta_{m_{\mathrm{C}}}\dot{m}_{\mathrm{C}} + \eta_{\mathrm{CC}}H_{\mathrm{u}}\dot{m}_{\mathrm{f}} - c_{\mathrm{p}}T_{\mathrm{CC}}\theta_{m_{\mathrm{T}}}\dot{m}_{\mathrm{T}}) - c_{\mathrm{v}}T_{\mathrm{CC}}(\theta_{m_{\mathrm{C}}}\dot{m}_{\mathrm{C}} + \dot{m}_{\mathrm{f}} - \theta_{m_{\mathrm{T}}}\dot{m}_{\mathrm{T}})]\\
&\quad + \frac{\gamma \mathrm{R} T_{\mathrm{CC}}}{V_{\mathrm{CC}}}(\theta_{m_{\mathrm{C}}}\dot{m}_{\mathrm{C}} + \dot{m}_{\mathrm{f}} - \theta_{m_{\mathrm{T}}}\dot{m}_{\mathrm{T}}),\\
\dot{P}_{\mathrm{NLT}} &= \frac{T_{\mathrm{M}}}{V_{\mathrm{M}}}(\theta_{m_{\mathrm{T}}}\dot{m}_{\mathrm{T}} + \frac{\beta}{\beta+1}\theta_{m_{\mathrm{C}}}\dot{m}_{\mathrm{C}} - \dot{m}_{\mathrm{Nozzle}}),
\end{aligned} \tag{24}$$

For the physical significance of the model parameters and details refer to [29]. The five gas turbine measured outputs are considered to be the output pressure and temperature of the

compressor, the turbine, and the spool speed, namely

$$
\begin{aligned}
&y_1(t) = T_{\mathrm{C}} = T_{\mathrm{diffuser}}[1 + \frac{1}{\theta_{\eta_{\mathrm{C}}}\eta_{\mathrm{C}}}[(\frac{P_{\mathrm{CC}}}{P_{\mathrm{diffuzer}}})^{\frac{\gamma-1}{\gamma}} - 1]],\\
&y_2(t) = P_{\mathrm{CC}},\ y_3(t) = N,\ y_4(t) = P_{\mathrm{NLT}},\\
&y_5(t) = T_{\mathrm{CC}}[1 - \theta_{\eta_{\mathrm{T}}}\eta_{\mathrm{T}}(1 - (\frac{P_{\mathrm{NLT}}}{P_{\mathrm{CC}}})^{\frac{\gamma-1}{\gamma}}].
\end{aligned}
\tag{25}
$$

In order to discretize the above model for implementation of our proposed dual state/parameter estimation particle filters, a simple Euler Backward method is applied with the sampling period of $T_s = 10$ msec.

The system health parameters are represented by the compressor and the turbine efficiency, η_{C} and η_{T}, respectively, and the mass flow capacities, $\dot{m}_{\mathrm{C}}$ and $\dot{m}_{\mathrm{T}}$, respectively. A fault vector is incorporated in the above model to represent the effects of the system health parameters which are denoted by $\theta = [\theta_{\eta_{\mathrm{C}}}, \theta_{m_{\mathrm{C}}}, \theta_{\eta_{\mathrm{T}}}, \theta_{m_{\mathrm{T}}}]^{\mathrm{T}}$. Each parameter variation can be a manifestation of changes in the fault vector and which is considered as a multiplicative fault type. All the simulations are conducted in the cruise steady-state flight condition mode in which two of the system health parameters are considered to be *slowly time-varying*.

In order to demonstrate the effectiveness and capabilities of our proposed algorithms, the simulation results corresponding to the conventional kernel smoothing method [17] (that is obtained without using the RPE term in the parameter estimation algorithm) as well as a method that uses stochastic optimization with the step size as a decreasing series of time [16] are also provided for performance comparisons. In our work, the adaptive step size ($P_t^{(j)} = \gamma_t R_t^{(j)}$) is defined as the multiplication of a constant γ_t with $R_t^{(j)}$ which is evaluated on-line based on the trace of the covariance matrix of the prediction error which is estimated from the maximum likelihood method. The residuals corresponding to the parameter estimates are obtained. Based on the percentage of the maximum absolute error criterion, a convergence time of 2 seconds is obtained in the simulation results for estimating both the states and the parameters corresponding to 25 runs of simultaneous faults with severities ranging from 1% to 10%.

To choose the number of particles for implementation of the state and parameter estimation filters, a quantitative study was conducted. Specifically, based on the mean absolute error (MAE%) obtained in the estimation steady state operation and by considering the algorithm implementation computational time, the number of particles is chosen as $N = M = 50$ for both

state and parameter estimation filters in this application. Consequently, acceptable performance and convergence times are obtained. The shrinkage coefficient is also selected as 0.93. The initial distributions (mean and covariance matrices) of the states and parameters are selected to correspond to the cruise flight operational conditions as provided in [29].

The input fuel flow to the engine is changed (that is, it is decreased by 2%) one second after reaching the steady state condition. Afterwards, a simultaneous fault in all the 4 health parameters of the engine is applied at $t = 9$ sec. The fault for the compressor and the turbine efficiencies follow the pattern of a drift fault which starts at $t = 9$ sec and causes a 5% loss of effectiveness in the compressor efficiency by the end of the simulation time (i.e. at $t = 19$ sec), and a 3% loss of effectiveness in the turbine efficiency by the end time of the simulation time. Simultaneously, the mass flow capacities of both the compressor and the turbine are affected by a fault which causes a 5% loss of effectiveness.

The dual estimation results corresponding to the above three methods are provided in Figures 1 and 2 for the estimates of the states and health parameters, respectively. The presented simulation results show that our proposed method in terms of its convergence time and the MAE$\%$ outperforms the other two methods significantly.

For the method based on the stochastic optimization with a decreasing step size, the step size after <u>extensive</u> trial and error is chosen as $\frac{0.85}{(k+3)^{0.602}}$ for this application, which yields a competition with our proposed adaptive step size method. However, it should be emphasized that one is generally interested in implementing an algorithm that is capable of tracking the parameter changes <u>under all conditions</u> and without the need for extensive fine tuning of the step size in each scenario and for each specific application. Nevertheless, the accuracy of the algorithm in [16] after fault occurrence, specifically in the case of slow time-varying changes in the system health parameters, degrades considerably.

The residuals related to the compressor mass flow faults and turbine efficiency faults are shown in Figure 3. These results show that in case of changes in the engine input (at $t = 1$ sec) the selected residuals do not exceed their confidence bounds, and the results in Table I(a) show that the maximum MAE$\%$ for both the state and parameter estimates in this scenario are between $0.1\% - 0.5\%$ of their nominal values. However, in the worst case the post fault estimated MAE of m_{T} is 0.47% of its nominal value. The percentage of the MAE measurement (output) estimate error as given in Table I(b) shows that after simultaneous fault occurrences the error increases

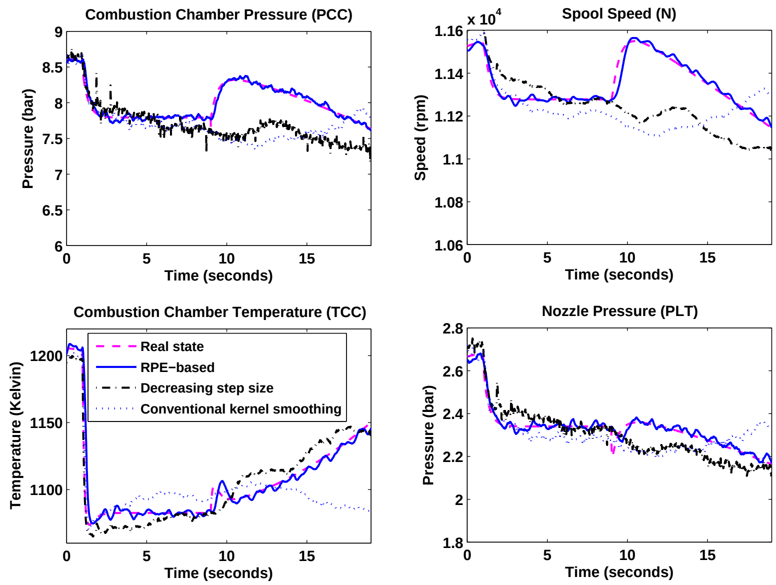

Fig. 1. Estimated states for the simultaneous faults using RPE-based (this work), Decreasing step size [16], and Conventional kernel smoothing [17] approaches.

TABLE I

(A)STATE/PARAMETER, (B)OUTPUT ESTIMATION PERCENTAGE OF MAES IN CASE OF SIMULTANEOUS FAULT SCENARIOS

(a)

State	Before Fault	After Fault
$P_{\rm CC}$	0.2217	0.2372
N	0.0535	0.1061
$T_{\rm CC}$	0.2086	0.1928
$P_{\rm NLT}$	0.3970	0.4291
$\eta_{\rm C}$	0.1735	0.1821
$m_{\rm C}$	0.2811	0.3293
$\eta_{\rm T}$	0.1016	0.1485
$m_{\rm T}$	0.4589	0.4744

(b)

Output	Before Fault	After Fault
$T_{\rm C}$	0.2207	0.2852
$P_{\rm C}$	1.2926	1.3729
N	0.0535	0.1027
$T_{\rm T}$	0.1565	0.1650
$P_{\rm T}$	2.1210	2.2348

when compared to their values before the fault occurrence. This is caused due to the accumulation of parameter estimation errors while all the four parameters are affected by a fault.

To summarize, our proposed fault diagnosis algorithm is capable of detecting, isolating and estimating the component faults of a gas turbine engine with an accuracy of 0.3% for the compressor and 0.5% for the turbine faults.

Fault Diagnosis Confusion Matrix Analysis

Finally, in this subsection a quantitative study is performed by utilizing the confusion matrix analysis [38] to evaluate the increase in the rate of the false alarms and/or misclassifications of faults in our considered application when the detection thresholds are fixed at the values

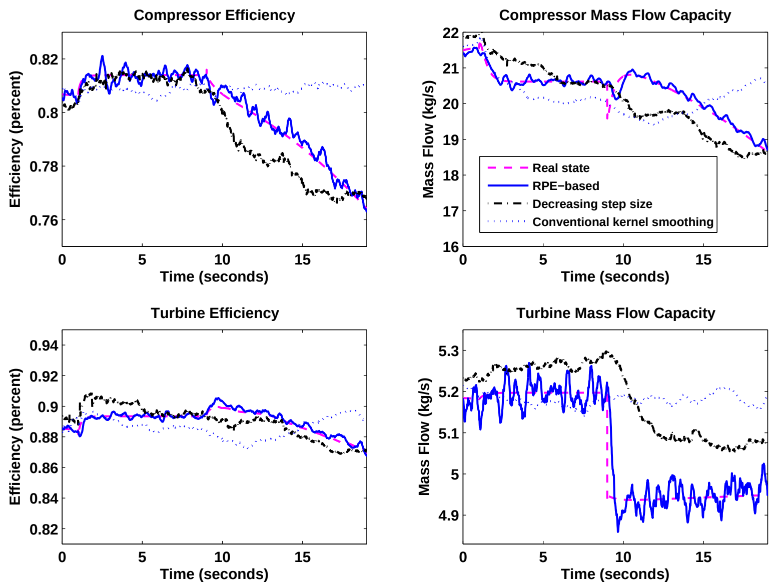

Fig. 2. Estimated health parameters for the simultaneous faults using RPE-based (this work), Decreasing step size [16], and Conventional kernel smoothing [17] approaches.

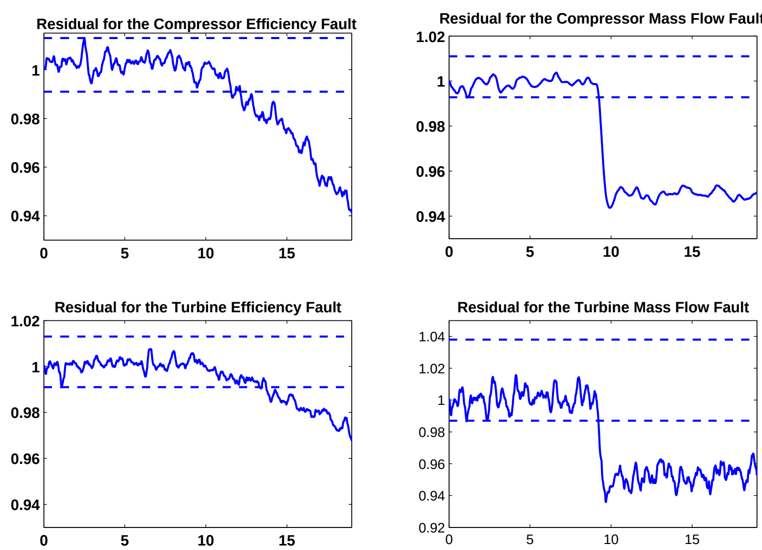

Fig. 3. Residuals related to the simultaneous fault scenarios.

that are obtained from the 25 runs of the simultaneous fault scenarios shown in Figure 3 (for the compressor mass flow fault and the turbine efficiency fault parameters). The confusion matrix data is obtained by performing Monte Carlo simulations for another 35 simultaneous fault scenarios with different fault severities and in presence of process and measurement noise having 4 different covariance levels ranging from 50% to 120% of the nominal values of the

TABLE II

CONFUSION MATRIX FOR (A) 50% OF THE NOMINAL NOISE COVARIANCE, (B) 120% OF THE NOMINAL NOISE COVARIANCE

(a)

Fault	η_{C}	m_{C}	η_{T}	m_{T}	NoFault
η_{C}	**31**	0	2	2	0
m_{C}	0	**30**	2	3	0
η_{T}	1	1	**28**	4	1
m_{T}	1	1	3	**29**	1
NoFault	0	0	1	1	**33**

(b)

Fault	η_{C}	m_{C}	η_{T}	m_{T}	NoFault
η_{C}	**29**	0	2	4	0
m_{C}	0	**29**	3	3	0
η_{T}	1	1	**27**	3	3
m_{T}	1	1	4	**26**	3
NoFault	0	0	2	3	**30**

process and measurement noise covariances (according to [29]). In these scenarios, at each time more than one of the system health parameters are affected by component faults.

The results are presented in Tables II(a) and II(b) in which the rows depict the actual number of fault categories applied and the columns represent the number of estimated fault categories. The diagonal elements represent the true positive rate (TP) for each fault occurrence case. The accuracy (AC), precision (P) and the false positive rate (FP) of our proposed algorithm are also evaluated from the results of the confusion matrix analysis according to the following formulae [38],

$$AC = \frac{\sum_{j=1}^{5} c_{jj}}{\sum_{i=1}^{5}\sum_{j=1}^{5} c_{ij}},\ P_j = \frac{c_{jj}}{\sum_{i=1}^{5} c_{ij}},\ FP = \frac{\sum_{j=1}^{4} c_{5j}}{\sum_{j=1}^{5} c_{5j}},$$

where $c_{ij},\ i,j = 1,...,5$ denote the elements of the confusion matrix. In Table III the results of the confusion matrix analysis according to the above metrics for both Tables II(a) and II(b) are provided. The results show that as the noise covariances increase from 50% to 120% of their nominal values, the accuracy of the fault diagnosis algorithm decreases by 5.72%, and the false positive rate increases by 8.58%. The precision of the algorithm remains robust for diagnosis of the compressor health parameters, the precision decreases by 6.73% for the fault diagnosis of the turbine efficiency and also decreases by 7.69% for the fault diagnosis of the turbine mass flow capacity.

TABLE III

CONFUSION MATRIX ANALYSIS RESULTS

Noise Level	$AC\%$	$FP\%$	$P_{\eta_{\mathrm{C}}}\%$	$P_{m_{\mathrm{C}}}\%$	$P_{\eta_{\mathrm{T}}}\%$	$P_{m_{\mathrm{T}}}\%$
50% of Nom. Cov.	86.29	5.71	93.94	93.75	77.78	74.36
120% of Nom. Cov.	80.57	14.29	93.55	93.55	71.05	66.67

V. Conclusion

In this paper, a novel dual estimation filtering scheme is proposed and developed based on particle filters to estimate a nonlinear stochastic system states and time variations in its parameters. A dual structure is proposed for achieving simultaneous state and parameter estimation objectives. Performance results of the application of our method to a gas turbine engine under healthy and faulty scenarios are also presented to demonstrate and illustrate the superior capability and performance of our scheme for a challenging fault diagnostic application as compared to two other alternative techniques that are available in the literature. The estimation results accuracy in terms of the fault identification are also provided. The obtained results are validated by performing a confusion matrix analysis.

References

[1] J. Korbicz, Z. Fathi, and W. Fred Ramirez, "State estimation schemes for fault detection and diagnosis in dynamic systems," *International Journal of Systems Science*, vol. 24, no. 5, pp. 985–1000, 1993.

[2] H. Kushner and H. Joseph, *Probability methods for approximations in stochastic control and for elliptic equations*, vol. 129. Academic Press New York, 1977.

[3] V. Beneš, "Exact finite-dimensional filters for certain diffusions with nonlinear drift," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 5, pp. 65–92, 1981.

[4] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.

[5] A. Bensoussan, R. Glowinski, and A. Raşcanu, "Approximation of some stochastic differential equations by the splitting up method," *Applied Mathematics & Optimization*, vol. 25, pp. 81–106, 1992.

[6] S. Lototsky and B. Rozovskii, "Recursive multiple wiener integral expansion for nonlinear filtering of diffusion processes," *Stochastic Processes and Functional Analysis, Lecture Notes in Pure and Appl. Math*, vol. 186, pp. 199–208, 1997.

[7] D. Crisan, J. Gaines, and T. Lyons, "Convergence of a branching particle method to the solution of the zakai equation," *SIAM Journal on Applied Mathematics*, vol. 58, pp. 1568–1590, 1998.

[8] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, 2009.

[9] E. Laroche, E. Sedda, and C. Durieu, "Methodological insights for online estimation of induction motor parameters," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 5, pp. 1021–1028, 2008.

[10] R. Li, M. A. Henson, and M. J. Kurtz, "Selection of model parameters for off-line parameter estimation," *IEEE Transactions on Control Systems Technology*, vol. 12, no. 3, pp. 402–412, 2004.

[11] N. Tudoroiu and K. Khorasani, "Satellite fault diagnosis using a bank of interacting kalman filters," *IEEE Transactions on Aerospace and Electronic Systems,*, vol. 43, no. 4, pp. 1334–1350, 2007.

[12] E. A. Wan, R. Van Der Merwe, and A. T. Nelson, "Dual estimation and the unscented transformation," in *Advances in Neural Information Processing Systems 12*, pp. 666–672, MIT Press, 1999.

[13] H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser, "Dual state–parameter estimation of hydrological models using ensemble Kalman filter," *Advances in Water Resources*, vol. 28, no. 2, pp. 135–147, 2005.

[14] N. Daroogheh, N. Meskin, and K. Khorasani, "Particle filtering for state and parameter estimation in gas turbine engine fault diagnostics," in *Proceedings of the American Control Conference*, 2013.

[15] N. Kantas, A. Doucet, S. S. Singh, and J. M. Maciejowski, "An overview of sequential monte carlo methods for parameter estimation in general state-space models," in *15th IFAC Symposium on System Identification (SYSID), Saint-Malo, France.(invited paper)*, vol. 102, p. 117, 2009.

[16] X. Yang, K. Xing, K. Shi, and Q. Pan, "Joint state and parameter estimation in particle filtering and stochastic optimization," *Journal of Control Theory and Applications*, vol. 6, pp. 215–220, 2008.

[17] J. Liu and M. West, *Combined parameter and state estimation in simulation-based filtering*. Institute of Statistics and Decision Sciences, Duke University, 1999.

[18] M. West, "Mixture models, Monte Carlo, Bayesian updating, and dynamic models," *Computing Science and Statistics*, pp. 325–325, 1993.

[19] L. Ljung and T. Söderström, *Theory and Pactice of Recursive Identification*. 1983.

[20] I. Collings, V. Krishnamurthy, and J. Moore, "On-line identification of hidden Markov models via recursive prediction error techniques," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3535–3539, 1994.

[21] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2010.

[22] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems*. Springer Publishing Company, Incorporated, 2012.

[23] J. Dong, M. Verhaegen, and F. Gustafsson, "Robust fault detection with statistical uncertainty in identified parameters," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5064–5076, 2012.

[24] R. Isermann, *Fault-diagnosis systems*. Springer, 2006.

[25] M. E. Orchard, *A Particle Filtering-based Framework for On-line Fault Diagnosis and Failure Prognosis*. PhD thesis, Georgia Institute of Technology, 2006.

[26] G. Gonzalez, R. Paut, A. Cipriano, D. R. Miranda, and G. Ceballos, "Fault detection and isolation using concatenated wavelet transform variances and discriminant analysis," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1727–1736, 2006.

[27] L. Guo, Y.-M. Zhang, H. Wang, and J.-C. Fang, "Observer-based optimal fault detection and diagnosis using conditional probability distributions," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3712–3719, 2006.

[28] E. Sobhani-Tehrani and K. Khorasani, *Fault diagnosis of nonlinear systems using a hybrid approach*, vol. 383. Springer, 2009.

[29] N. Meskin, E. Naderi, and K. Khorasani, "A multiple model-based approach for fault diagnosis of jet engines," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 1, pp. 254–262, 2013.

[30] E. Naderi, N. Meskin, and K. Khorasani, "Nonlinear fault diagnosis of jet engines by using a multiple model-based approach," *Journal of Engineering for Gas Turbines and Power*, vol. 134, no. 1, p. 011602, 2012.

[31] X.-L. Hu, T. Schon, and L. Ljung, "A general convergence result for particle filtering," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3424–3429, 2011.

[32] C. Musso, N. Oudjane, and F. LeGland, "Improving regularised particle filters," *Sequential Monte Carlo methods in practice*, pp. 247–271, 2001.

[33] A. Budhiraja, L. Chen, and C. Lee, "A survey of numerical methods for nonlinear filtering problems," *Physica D: Nonlinear Phenomena*, vol. 230, no. 1, pp. 27–36, 2007.

[34] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, “A review of process fault detection and diagnosis: Part i: Quantitative model-based methods,” *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.

[35] R. Isermann, “Process fault detection based on modeling and estimation methodsa survey,” *Automatica*, vol. 20, no. 4, pp. 387–404, 1984.

[36] D. Wang, M. Yu, C. B. Low, and S. Arogeti, *Model-based Health Monitoring of Hybrid Systems*. Springer, 2013.

[37] S. Simani, *Model-based fault diagnosis in dynamic systems using identification techniques*. PhD thesis, dell’Universit di Ferrara, 2003.

[38] R. Kohavi and F. Provost, “Confusion matrix,” *Machine Learning*, vol. 30, no. 2-3, pp. 271–274, 1998.